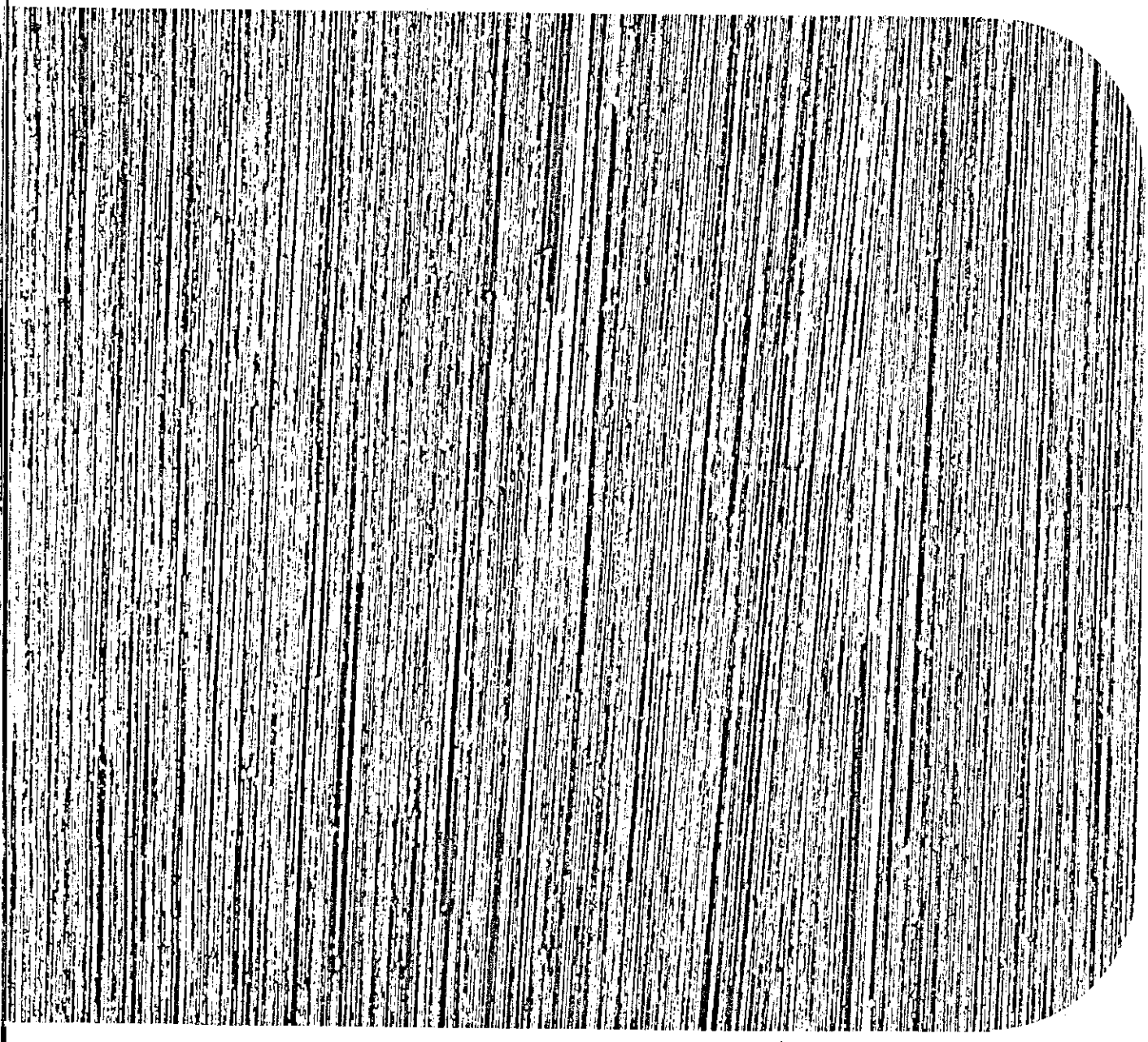
Australian Bureau of Statistics

HOUSEHOLD EXPENDITURE SURVEY 1974-75



HOUSEHOLD EXPENDITURE SURVEY 1974-75

Bulletin 3

STANDARD ERRORS

Australian Bureau of Statistics Canberra, Australia

Catalogue No. 6509.0
(Previously Ref. No. 17.21)

INQUIRIES

If you want to know more about these statistics ring Mr Jack Colwell, Canberra 526303, or write to Information Services, ABS, P.O. Box 10, Belconnen, A.C.T. 2616.

For copies of this publication contact Information Services, Canberra 526627 or State offices.

PREFACE

The 1974-75 Household Expenditure Survey was the first major survey of this kind to be conducted by the Australian Bureau of Statistics. The survey was designed to find out how the expenditure pattern of private households varies according to income level and other characteristics such as size and composition of the household, age and occupational status of the household head.

The results of this survey are being published in a series of bulletins, several of which have already been released. For details of the publications in the series see page 63.

The bulletin provides some discussion of the sampling and non-sampling errors associated with the published results of the survey. Sampling errors have mainly been presented in graphical form but some standard error tables are also provided. Table 3.1 on page 9 provides a key to the location of the appropriate graph or table for estimating the standard error of any published survey estimate.

This bulletin has been compiled by staff of the Sampling Section of the Australian Bureau of Statistics.

R.J. Cameron
Australian Statistician

Australian Bureau of Statistics
CANBERRA A.C.T.
November 1977

CONTENTS

SECTION 1 – EXPLANATORY NOTES

1.1 Reliability of Estimates – Sampling and Non Sampling Errors

The statistics from the Household Expenditure Survey are subject to two types of errors which are discussed below.

Sampling Errors

Since the survey estimates are based on only a sample of households, it is likely that they will differ from the figures that would have been obtained from a collection based on all households using the same questionnaires and procedures. These differences are called sampling errors.

Non–Sampling Errors

These are all the other errors in the survey results, which occur for reasons other than sampling. Non–sampling errors include the following:

(a) errors due to non–response, ie because of differences in the characteristics and patterns of expenditure and income between respondent and non–respondent households.

(b) errors on the part of respondents and interviewers such as faulty wording of questions, misunderstanding of what is required, and inability or unwillingness to provide accurate information.

(c) errors in processing such as mistakes in the recording or coding of the data obtained.

(d) errors due to the collection of the data over an extended time period.

Non–sampling errors may be present in any collection, whether it be a complete collection from all households defined to be within the scope of the collection (ie the survey population), or one based on a sample drawn from the population. Non–sampling errors are difficult to measure in any collection. However, every effort was made to minimise their effects in the Household Expenditure Survey. For example, questionnaires were designed with great care and were extensively tested prior to the survey, interviewers were intensively trained, and call–backs undertaken to all initially non–responding households, in order to increase the final response rate. A more detailed discussion of non–sampling errors may be found in Appendix 1.

1.2 Measurement of Sampling Error – Standard Errors

The magnitude of the sampling error associated with a sample estimate depends on the following factors:

Sample Design

There are many different sample designs which could have been used to obtain a sample from which to collect data on household expenditure. The final design attempted to make the survey results as accurate as possible within cost and operational constraints. For details of this design see Appendix 3.

Sample Size

The larger the size of sample on which the estimate is based, the smaller will be the sampling error. However the larger the sample size the greater the cost, and therefore the final sample size must be a compromise between cost and accuracy.

Population Variability

The third factor which influences sampling error is the extent to which households differ in their expenditure on the particular item being measured. This is referred to as the 'Population Variability' for that item. The smaller the population variability of an item the more likely it is that the population will be well represented by the sample, and therefore the smaller the sampling error. Conversely, the more variable the item, the greater the sampling error. For example, expenditure on 'Food' can be expected to have a smaller population variability than expenditure on 'Household Equipment and Operation', and therefore, all other things being equal, will be subject to a smaller sampling error.

The sampling error associated with any estimate can be estimated from the sample results. The sample size is known and the effects of the sample design and the variability of the population can be estimated from the sample respondents. It should be noted that sampling errors estimated in this way are themselves survey estimates and are therefore also subject to sampling errors.

Two commonly used measures of sampling error are the standard error and the standard error percent.

Standard Error

The particular sample selected for the survey is only one of a large number of samples that could have

been selected using the same sample selection procedures. Each sample would yield a different set of estimates. The standard error measures the variation of all the possible sample estimates around the figure that would have been obtained from a comparable complete enumeration of the population. The standard error is therefore an estimate of the sampling error in absolute terms, eg the standard error of an estimate of 'Total Expenditure' is an estimate of the sampling error measured in terms of a dollar amount.

Standard Error Percent

The standard error percent is obtained by taking the standard error, dividing by the estimate, and multiplying by 100. The standard error percent measures the size of the sampling error relative to the size of the estimate, expressed as a percentage.

1.3 Interpretation of Standard Errors – Confidence Intervals

The standard error is a useful measure of the sampling error associated with an estimate, in that we can construct interval estimates with prescribed confidence that the interval includes the value which would have been obtained from an enumeration of all households in the survey population. Such an interval is referred to as a confidence interval and takes the following form. Given an estimate and the standard error of that estimate, there are about 2 chances in 3 that the complete enumeration value is within one standard error of the sample estimate and about 19 chances in 20 that it is within 2 standard errors. For example, the estimate of expenditure on 'Current Housing Costs' in Melbourne is $23.00 (See page 3, Bulletin 2) with a standard error of $0.53 (See page 21). This means we can be about 67% sure that the complete collection value lies within the range $22.47 – $23.53 and about 95% sure that it lies in the range $21.94 – $24.06. These ranges are called the 67% and 95% confidence intervals respectively.

The standard error percent is a relative measure of the sampling error and can be used to compare the degree of accuracy of estimates. The following example illustrates the use of the standard error percent. The estimate of expenditure on 'Current Housing Costs' in Melbourne is $23.00 with a standard error of $0.53, while the estimate of the number of households in Melbourne with weekly household income between $200 and $260 is 162,900 with a standard error of 6,400 households. It is difficult to compare the accuracy of these two figures as the standard error is measured in terms of dollars in one case and number of households in the other. The relative accuracies of the two figures can be assessed by comparing their standard error percents, ie 2.3 percent for current housing costs compared with 3.9 percent for number of households with income between $200 and $260.

To simplify presentation, sampling errors have been presented in this bulletin in the form of standard error percents. The standard error of an estimate is obtainable from the standard error percent of that estimate as follows:

$$\text{Standard error} = \frac{\text{Standard error percent} \times \text{estimate}}{100}$$

1.4 Method of Presentation

Although it would have been possible to produce and publish estimated standard errors or standard error percents for all the estimates presented in the Household Expenditure Survey bulletins, this was not done for a number of reasons:

(a) Presentation difficulties would have been encountered because of the large number of estimates involved.

(b) Computer production of all standard errors would have been very costly.

(c) The standard errors themselves are subject to sampling errors, and 'smoothing' or averaging them in the appropriate fashion was expected to improve their accuracy.

It was therefore decided to 'smooth' the standard errors. To assist with the presentation of these smoothed standard errors, published tables have been divided into five groups as follows:

Group I – Broad Expenditure (expenditure classified by 15 items) and Medium Expenditure (expenditure classified by 102 items).

Group II – Numbers of Households and Numbers of Persons.

GroupIII – Household and Personal Income by Source of Income.

GroupIV – Quantiles.

GroupV – Other Estimates.

For each group (except for Group V where actual standard errors are shown), a series of graphs has been produced showing standard error percent graphed against the sample size on which the estimate was based,

or against the estimate itself. Each line on a graph represents a particular item or a series of items in the group, and can be used to obtain the standard error percent for any estimate of that item (or series of items). For example, line A on Graph I(8) on page 23 can be used to estimate the standard error percent for all estimates of expenditure on 'Bread' in Group I tables, (ie in all capital cities, quarters, income ranges, etc).

Full details of the methods used to draw the graphs are given in Appendix 2. The graphs have been drawn on logarithmic graph scales. This was done for two main reasons. Firstly, the standard error lines were based on a logarithmic linear model and for this reason they are represented by a straight line on logarithmic graph paper, making the graphs easier to read. Secondly, logarithmic presentation allows a far greater range of values to be shown on the graphs without sacrificing the accuracy with which small figures can be read. The main grid lines have been drawn in heavily and their values have been marked. The fine lines between two adjacent broad lines divide the interval between the broad lines into smaller intervals each representing an equal number of units.

For example, in Graph I (1) (see page 16) on the 'Standard Error Percent' axis, the four grid lines between the points 7.5 and 10 represent the values 8.0, 8.5, 9.0, 9.5, while between the points 40 and 50, the four lines represent the values 42,44,46 and 48.

In Group V, graphing was not possible because of insufficient points to accurately draw standard error lines. For this group, tables of standard error percents have been given.

The standard error percents obtained from the graphs are not intended to give a precise measure of the sampling error for a particular estimate, only to provide an indication of the likely magnitude of the standard error percent for any estimate of the item or items to which the standard error line refers. As the standard errors themselves are subject to sampling errors, the standard error percent tables presented for Group V should also only be used as a guide to likely levels of standard error percents.

1.5 Key to Use of Standard Error Graphs and Tables

The graphs and tables in Section 2 of this bulletin can be used to obtain a standard error percent for each survey estimate published in any of the bulletins. For each group of survey estimates there is a discussion of the standard errors for that group, together with instructions and examples to assist in determining standard errors. Section 2.6 discusses the standard errors of derived statistics.

The procedure for obtaining an approximate standard error for a survey estimate involves the following steps:

(i) From Table 3.1 identify the appropriate standard error group for the bulletin and table containing the estimate. Refer to the page number given, for discussion and presentation of the standard errors for that group.

(ii) By reference to this discussion, identify the particular graph or table relevant to that estimate.

(iii) Determine the standard error percent for that estimate by following the steps set out for that group.

(iv) If required, convert the standard error percent to a standard error by dividing by 100 and multiplying by the estimate.

TABLE 3.1 KEY TO STANDARD ERROR GRAPHS

Bulletin	*Tables*	*Graph Group*	*Refer to page (of this bulletin)*
2	2.1 – 2.13	I	11
4	4.1 – 4.9	I	11
5	5.1 – 5.8	I	11
6	6.1 – 6.19	I	11
7	7.1 – 7.8	II	36
	7.9	V	46
	7.10 – 7.17	III	39
	7.18 – 7.26	IV	42
	7.27 – 7.31	II	36
	7.32 – 7.37	III	39
	7.38 – 7.43	IV	42
8	8.1 – 8.16	I	11
	8.17 – 8.22	IV	42
	8.23	V	46
	8.24 – 8.25	IV	42

SECTION 2 – STANDARD ERROR GRAPHS AND TABLES

2.1 GROUP I – BROAD AND MEDIUM EXPENDITURE TABLES

This group contains all tables which are presented in the broad expenditure format, ie Tables 2.1 to 2.13, 5.2 to 5.8, 6.3 to 6.19, 8.3 to 8.8, 8.10 and 8.12 to 8.16; and those tables presented in the medium expenditure format, ie Tables 4.1 to 4.9, 5.1, 6.1, 6.2, 8.1, 8.2, 8.9 and 8.11.

For this group standard errors are presented in the form of graphs of standard error percent against the number of households in the sample on which the estimate is based. The standard error percent for a particular estimate in one of these tables can be obtained as follows:

(i) For all tables apart from Table 4.9, refer to the table containing the estimate to find the sample number of households from the top line of the column containing the estimate. For Table 4.9 the sample numbers are: Sydney 2255, Melbourne 2544, Brisbane 1119, Adelaide 983, Perth 1089, Hobart 562, Canberra 543, All Capitals 9095.

(ii) Refer to Table 3.2 on page 12 of this bulletin where a list of broad and medium expenditure items can be found. Use this table to determine the graph and page number for the standard error of that estimate.

(iii) Look up the item on the key to the graph to determine the alphabetic code of the line corresponding to that item.

(iv) Read the standard error percent for that line from the graph using the sample size found in step (i).

Examples

Broad Expenditure

Table 2.8 of Bulletin 2 shows that average weekly expenditure on 'Food' for those households in Canberra with weekly household income between $140 and $200 is estimated to be $30.29. Having determined from Table 3.1 of this bulletin that Group I is the appropriate group, the standard error percent may be obtained as follows:

(i) The number of households in sample is found to be 81 by looking at the top row of the third column of Table 2.8.

(ii) Table 3.2 indicates that Graph I (8) on page 23 is the appropriate graph.

(iii) The key on this graph indicates that the 'Total Food' line is line F.

(iv) Looking up a sample size of 81 on this line gives the standard error percent as 5.2.

The standard error is therefore

$$\frac{5.2 \times \$30.29}{100} = \$1.57$$

The 67 percent confidence interval is
$\$30.29 - \$1.57$ to $\$30.29 + \$1.57 = \$28.72$ to $\$31.86$,

and the 95 percent confidence interval is
$\$30.29 - (2 \times \$1.57)$ to $\$30.29 + (2 \times \$1.57) = \$27.15$ to $\$33.43$.

Medium Expenditure

Table 4.1 of Bulletin 4 shows that the average expenditure on 'Electricity' for those households in'All Capital Cities' with weekly household income of less than $80, is estimated to be $1.53. Having determined from Table 3.1 of this bulletin that Group I is the appropriate group, the standard error percent may be obtained as follows:

(i) The sample size is found to be 1,347 households from the top line of the first column of Table 4.1.

(ii) Table 3.2 indicates that Graph I (7) on page 22 is the appropriate graph.

(iii) The key on the graph indicates that line A is the appropriate line.

(iv) Looking at a sample size of 1,347 on this line gives the standard error percent as 1.7.

The standard error is therefore

$$\frac{1.7 \times \$1.53}{100} = \$0.03$$

The 67 percent confidence interval is $1.50 to $1.56 and the 95 percent confidence interval is $1.47 to $1.59.

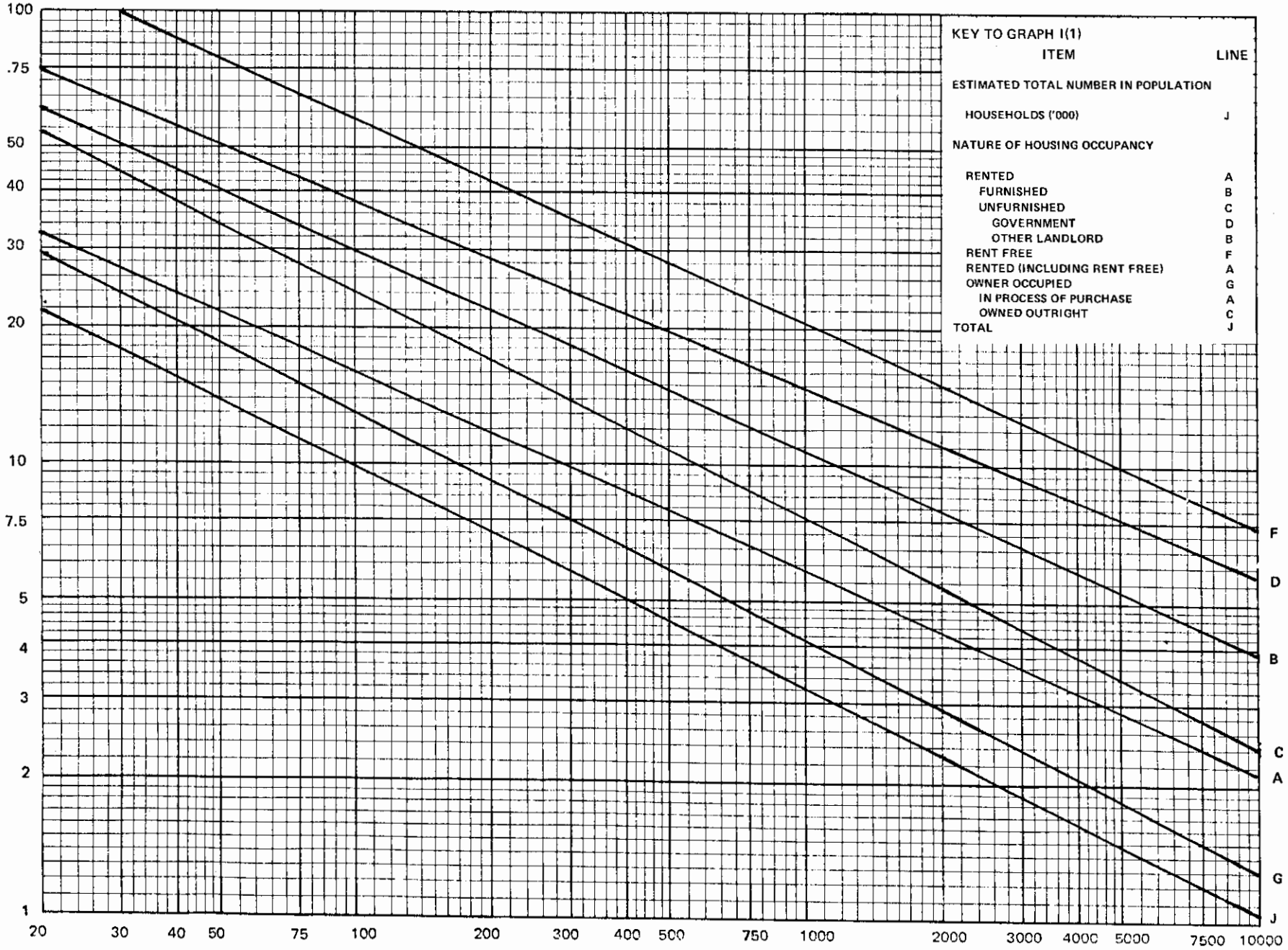
TABLE 3.2 GROUP I: BROAD AND MEDIUM EXPENDITURE ITEMS

TABLE 3.2 (CONT.) GROUP I: BROAD AND MEDIUM EXPENDITURE ITEMS

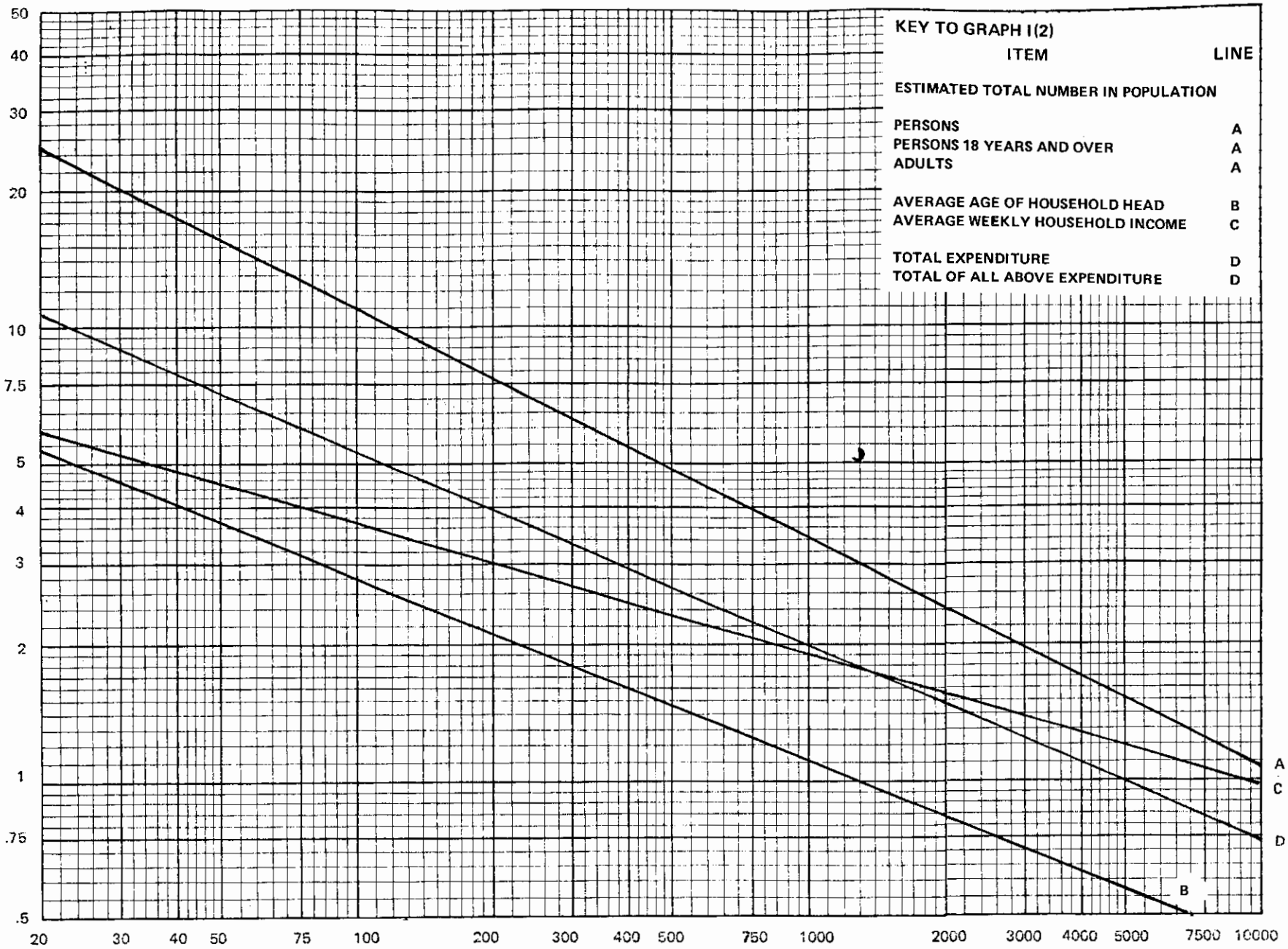
TABLE 3.2 (CONT.) GROUP I: BROAD AND MEDIUM EXPENDITURE ITEMS

TABLE 3.2 (CONT.) GROUP I: BROAD AND MEDIUM EXPENDITURE ITEMS

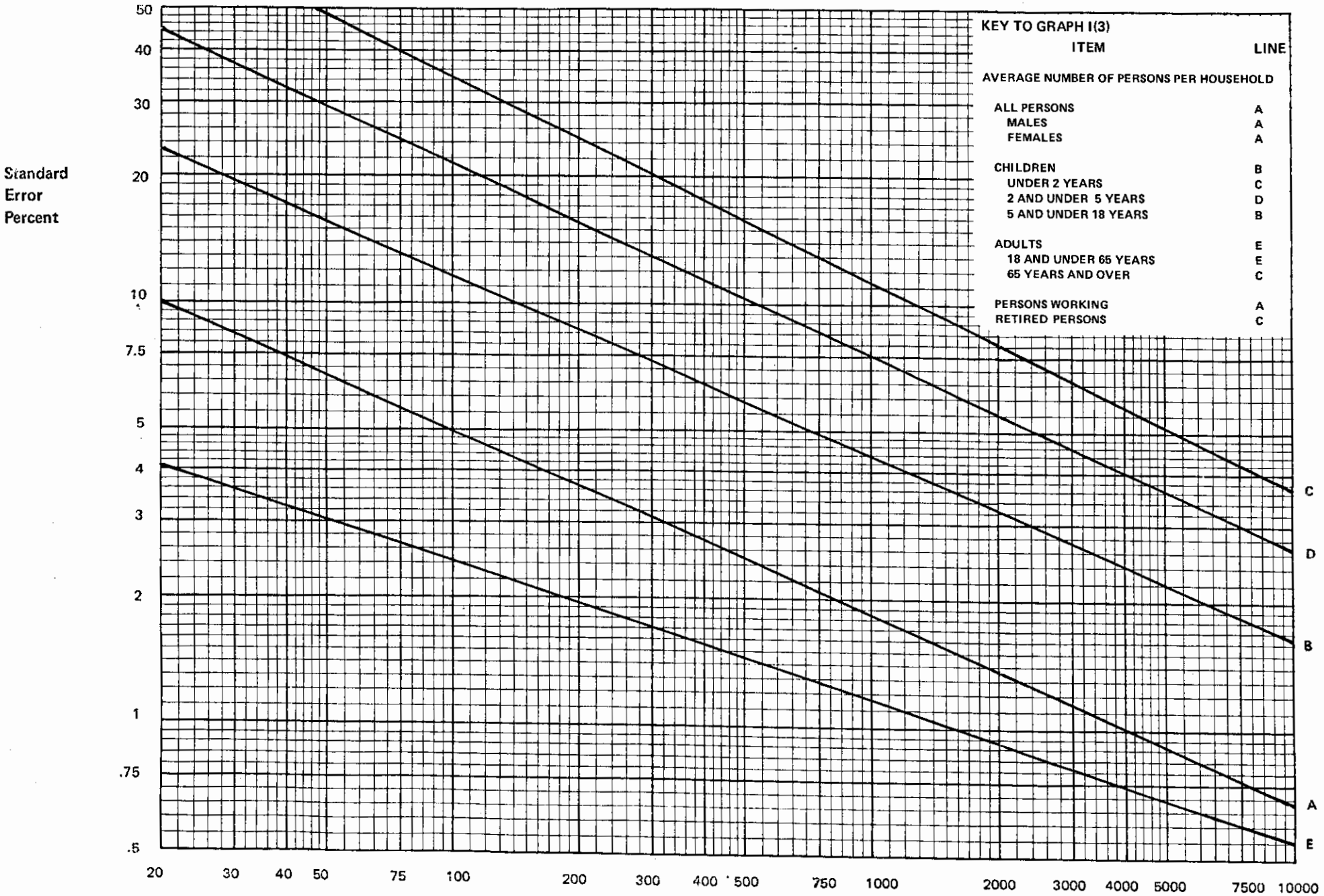
I(1) -- NATURE OF HOUSING OCCUPANCY -- NUMBER OF HOUSEHOLDS



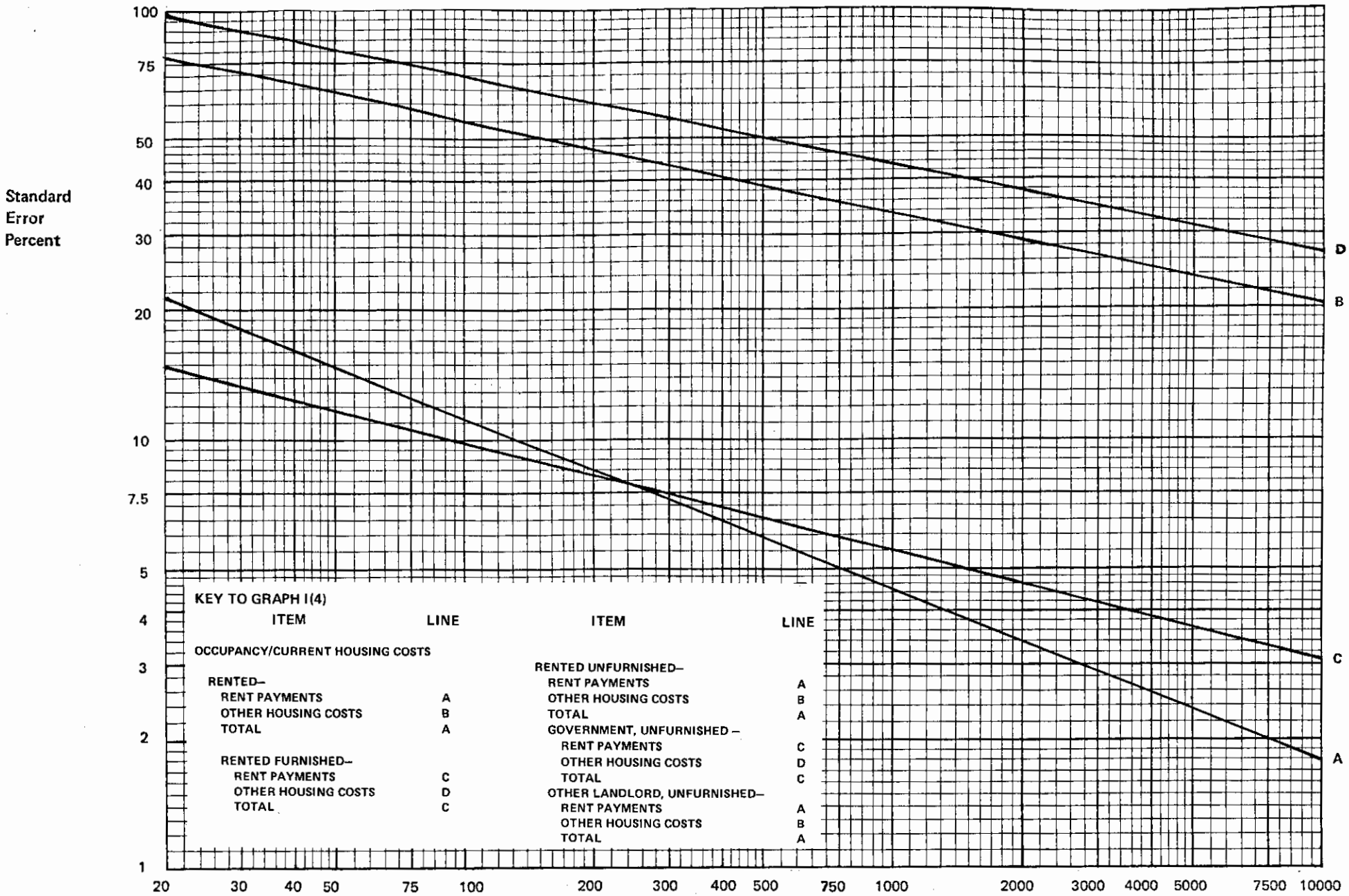
I(2) -- MISCELLANEOUS



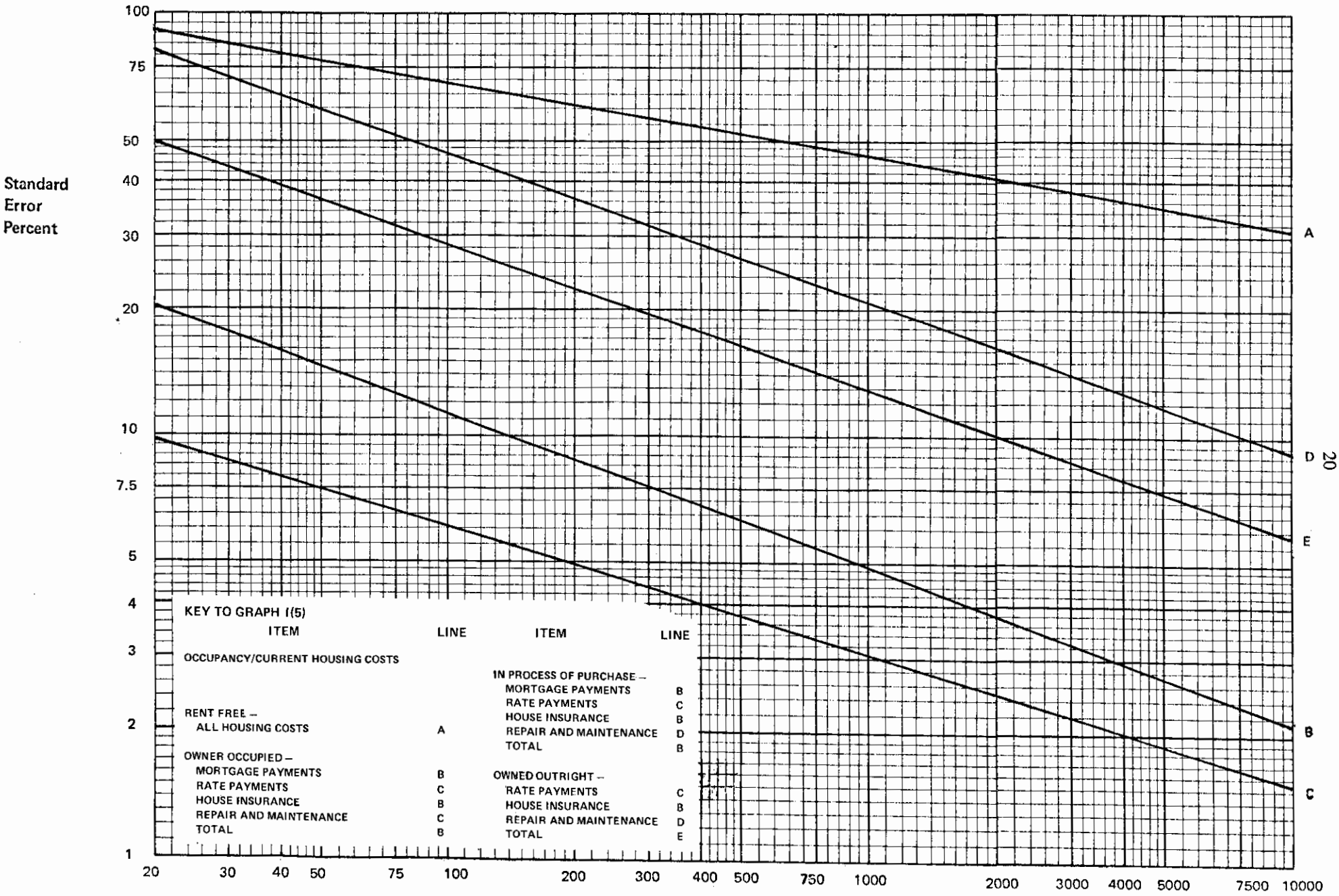
I(3) – AVERAGE NUMBER OF PERSONS PER HOUSEHOLD



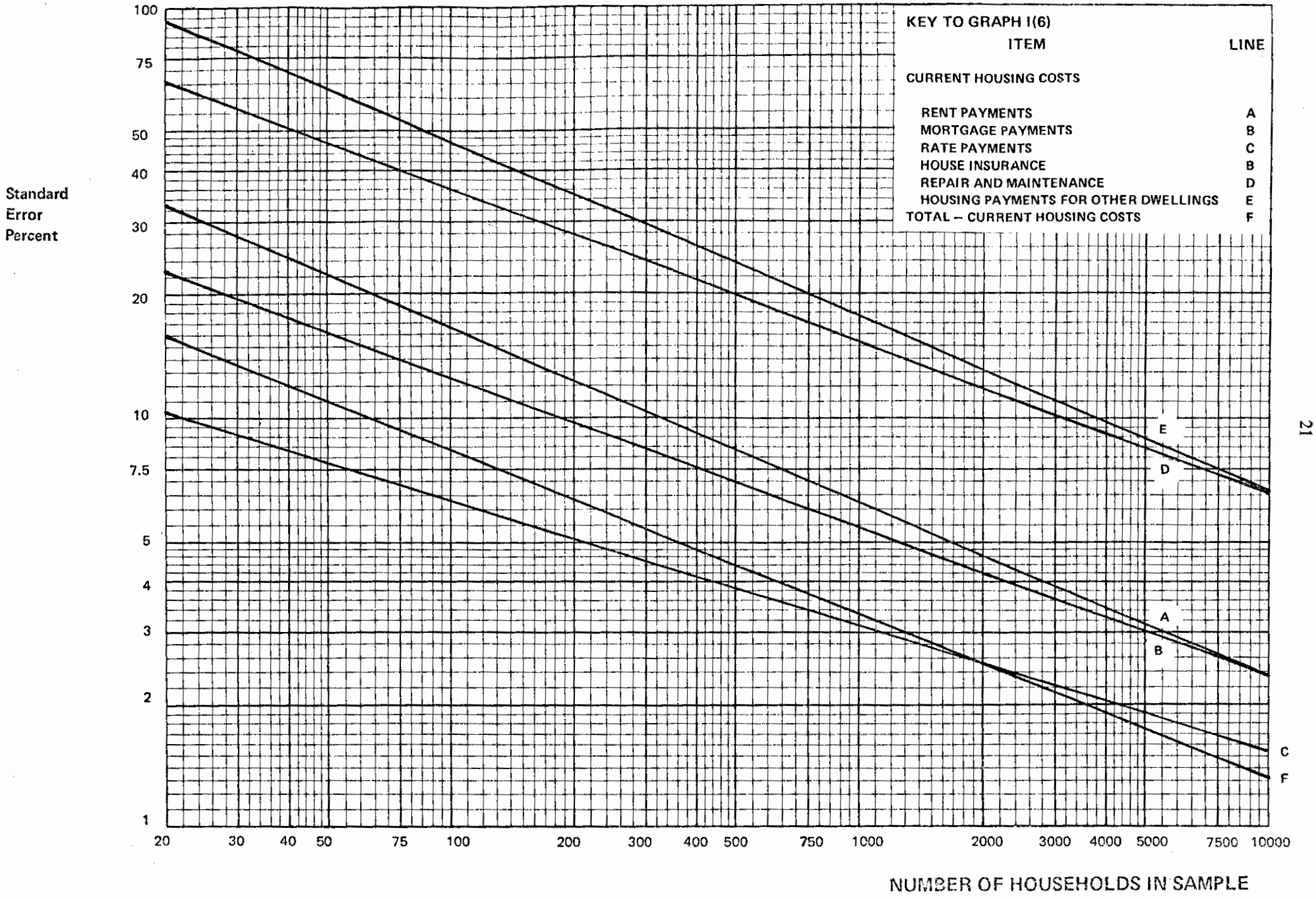
I(4) – OCCUPANCY/CURRENT HOUSING COSTS – RENTED



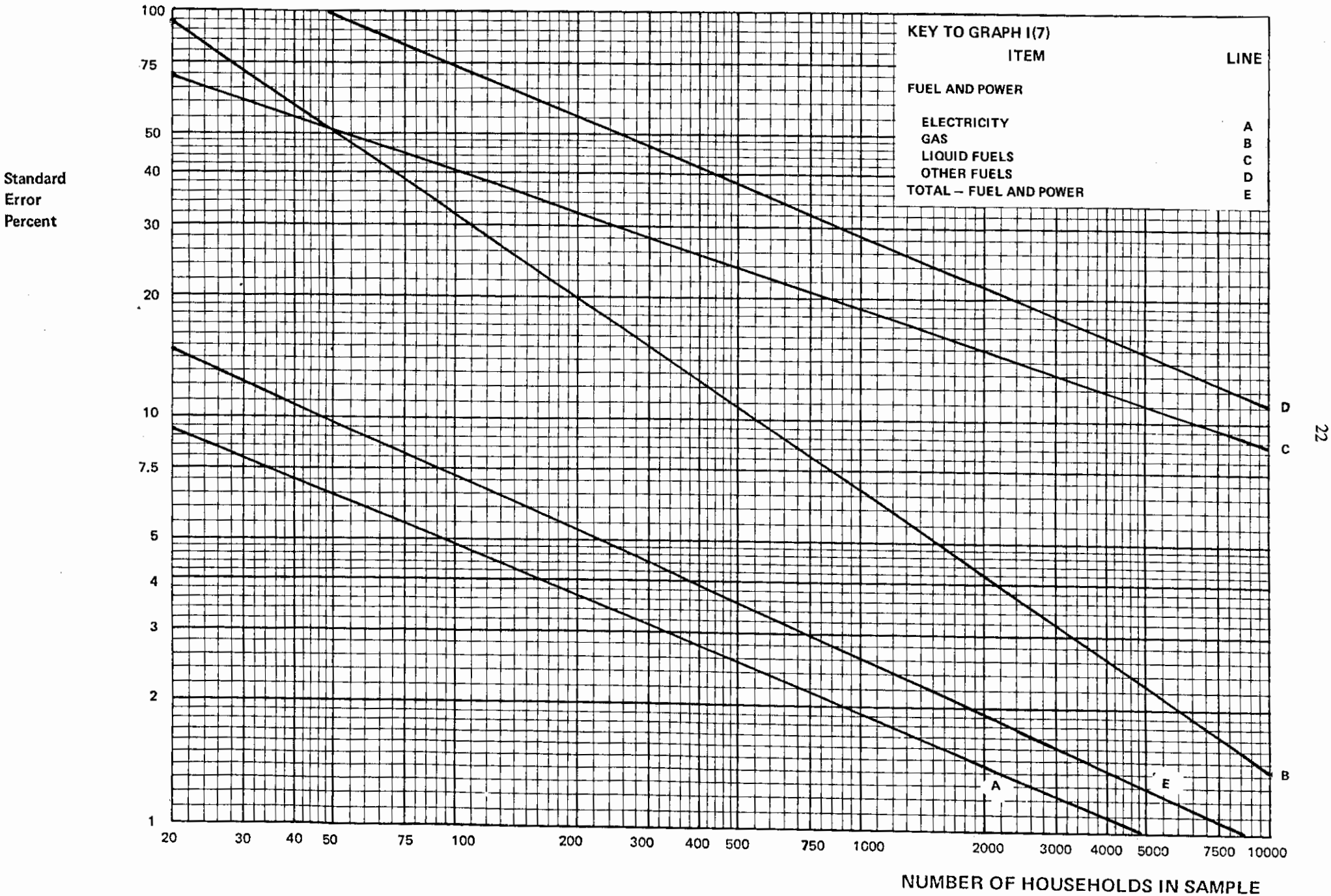
I(5) -- OCCUPANCY/CURRENT HOUSING COSTS -- RENT FREE AND OWNER OCCUPIED



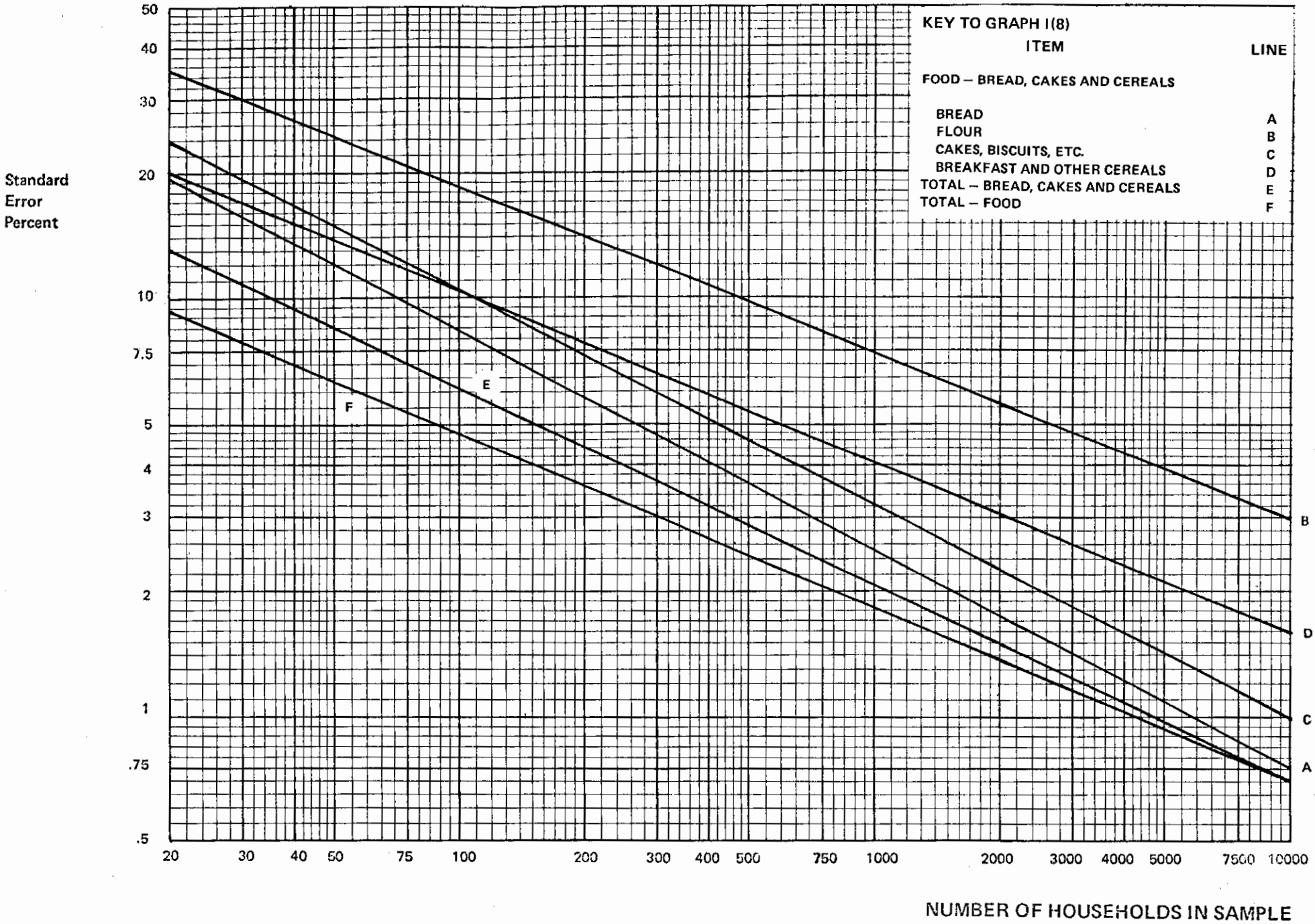
I(6) – EXPENDITURE ON CURRENT HOUSING COSTS



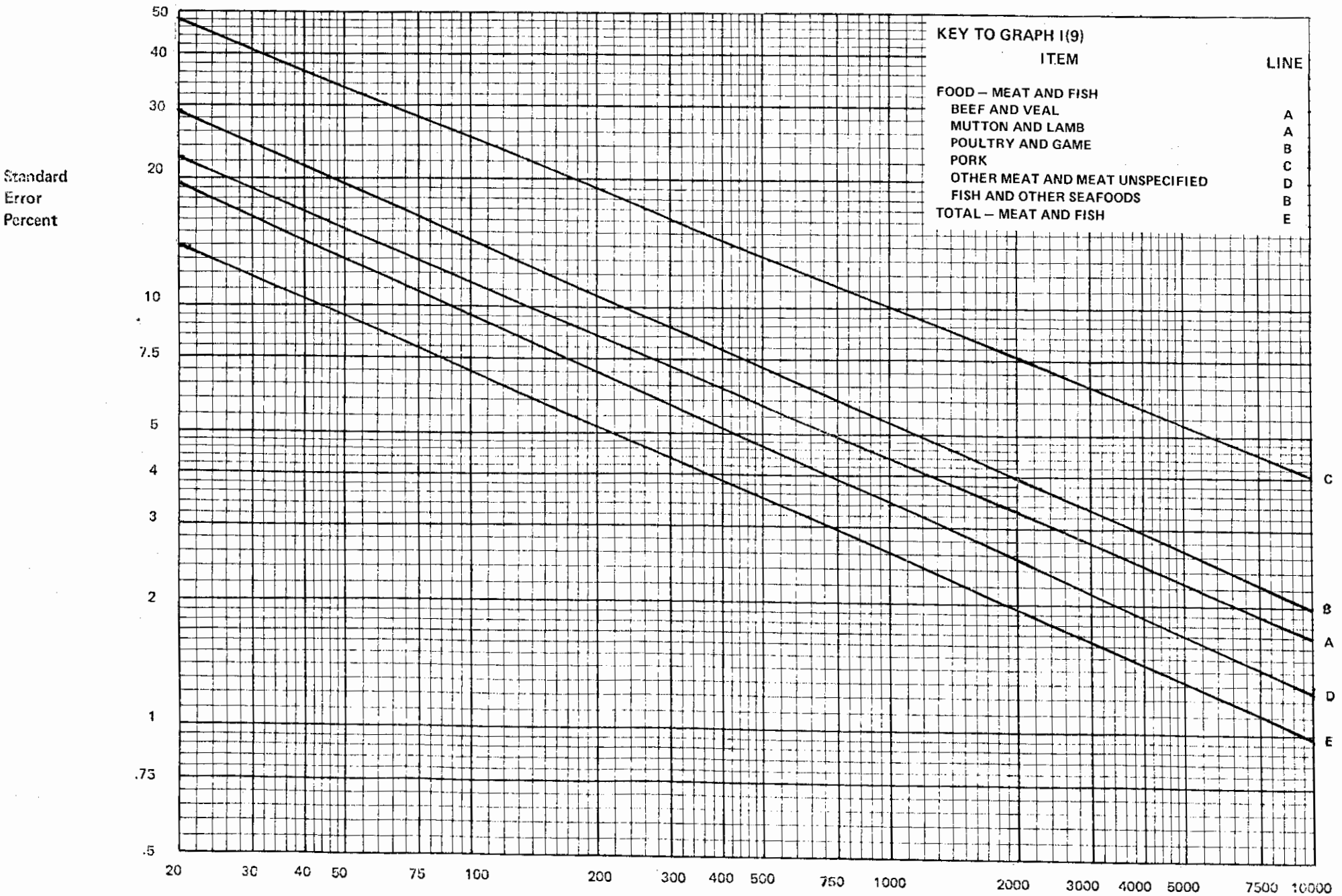
I(7) – EXPENDITURE ON FUEL AND POWER



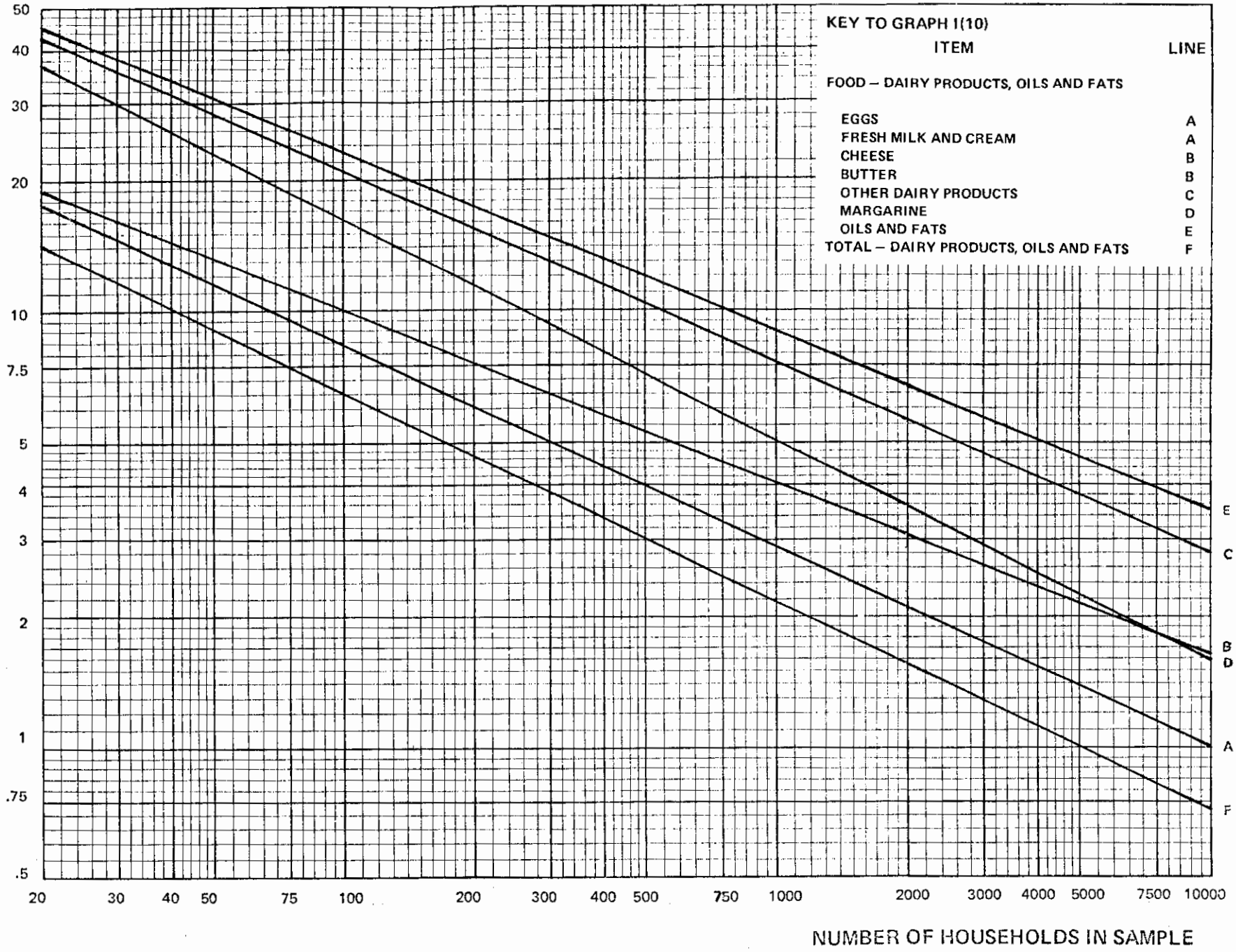
I(8) – EXPENDITURE ON BREAD, CAKES AND CEREALS



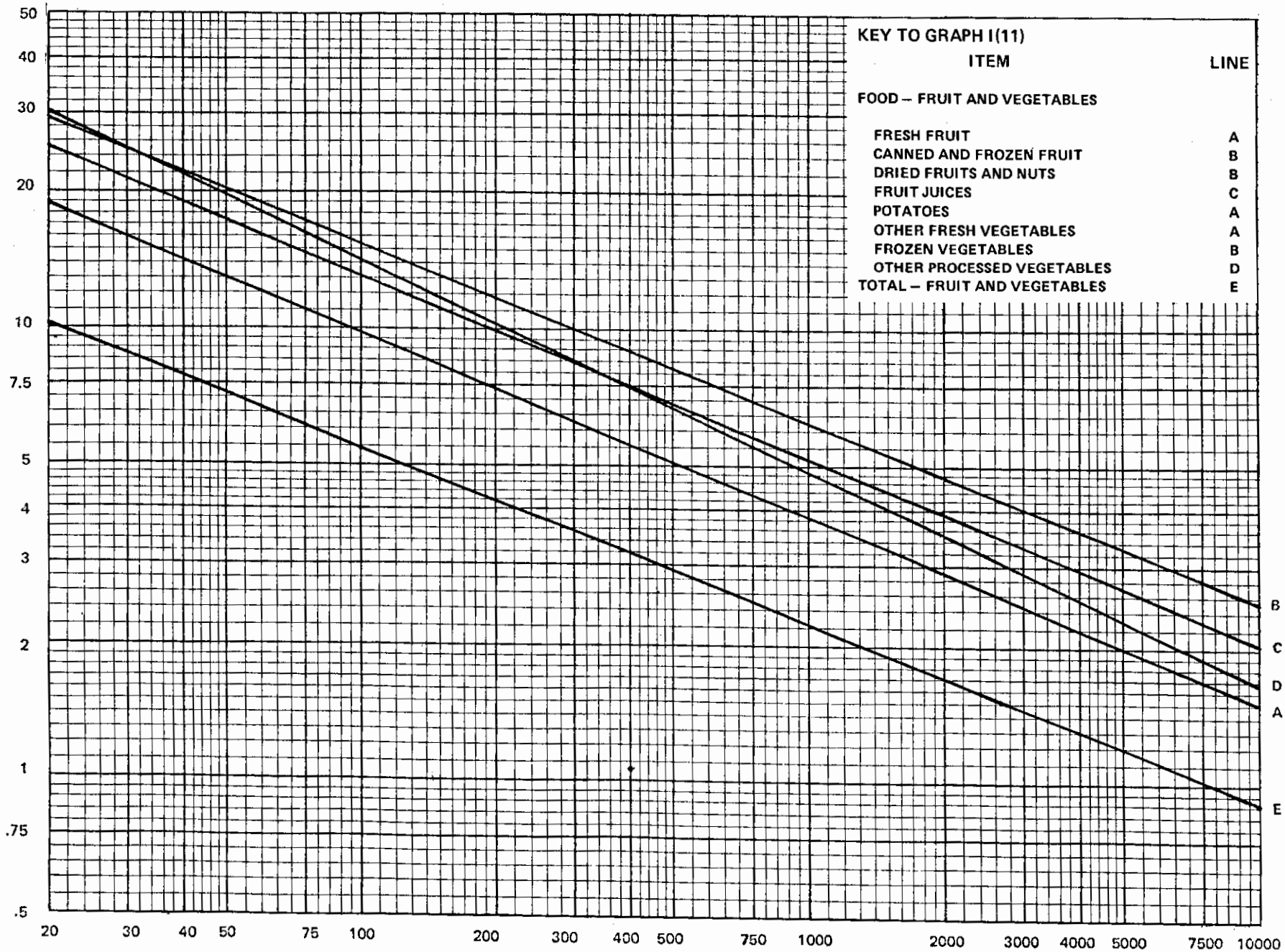
I(9) – EXPENDITURE ON MEAT AND FISH



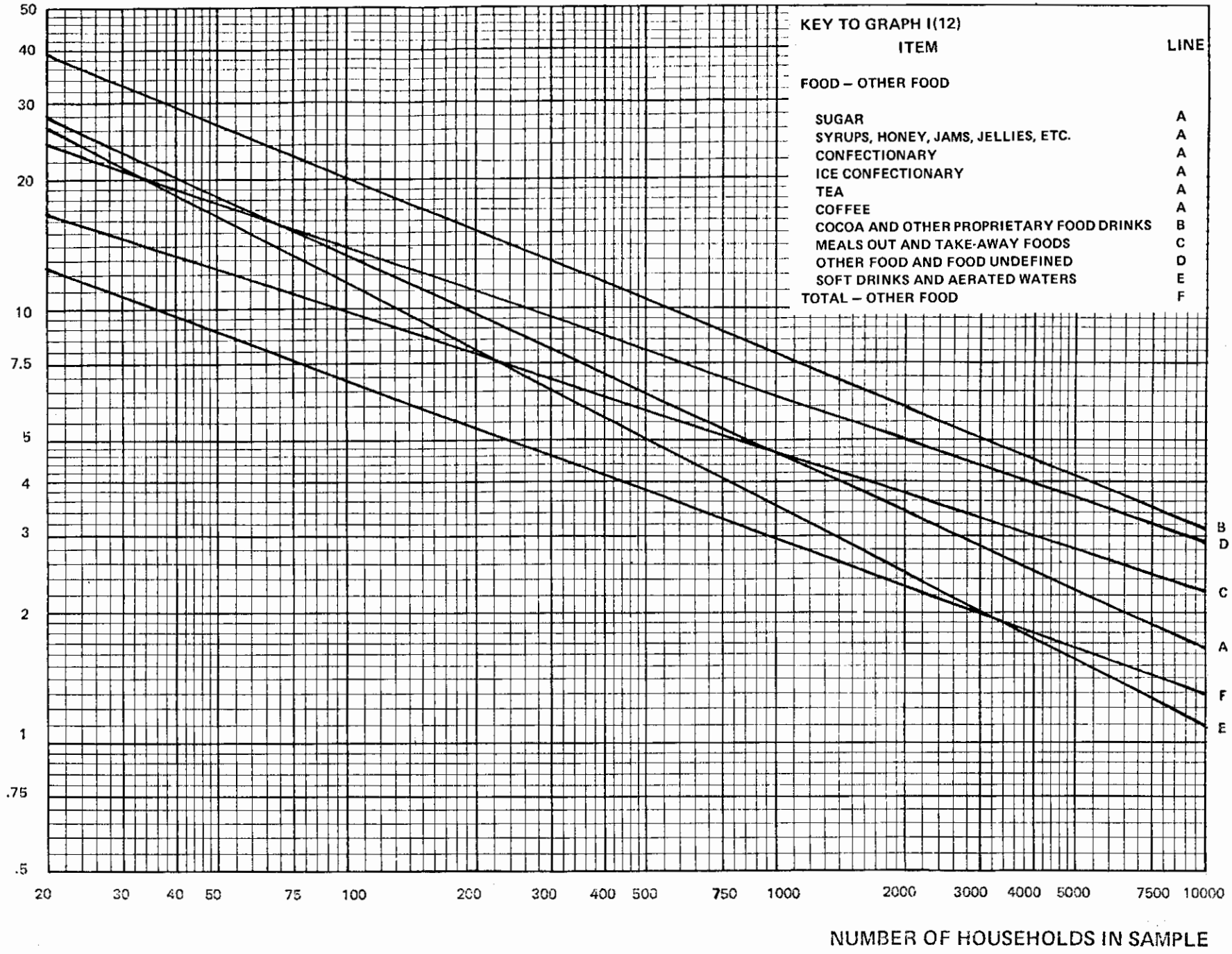
I(10) – EXPENDITURE ON DAIRY PRODUCTS, OILS AND FATS



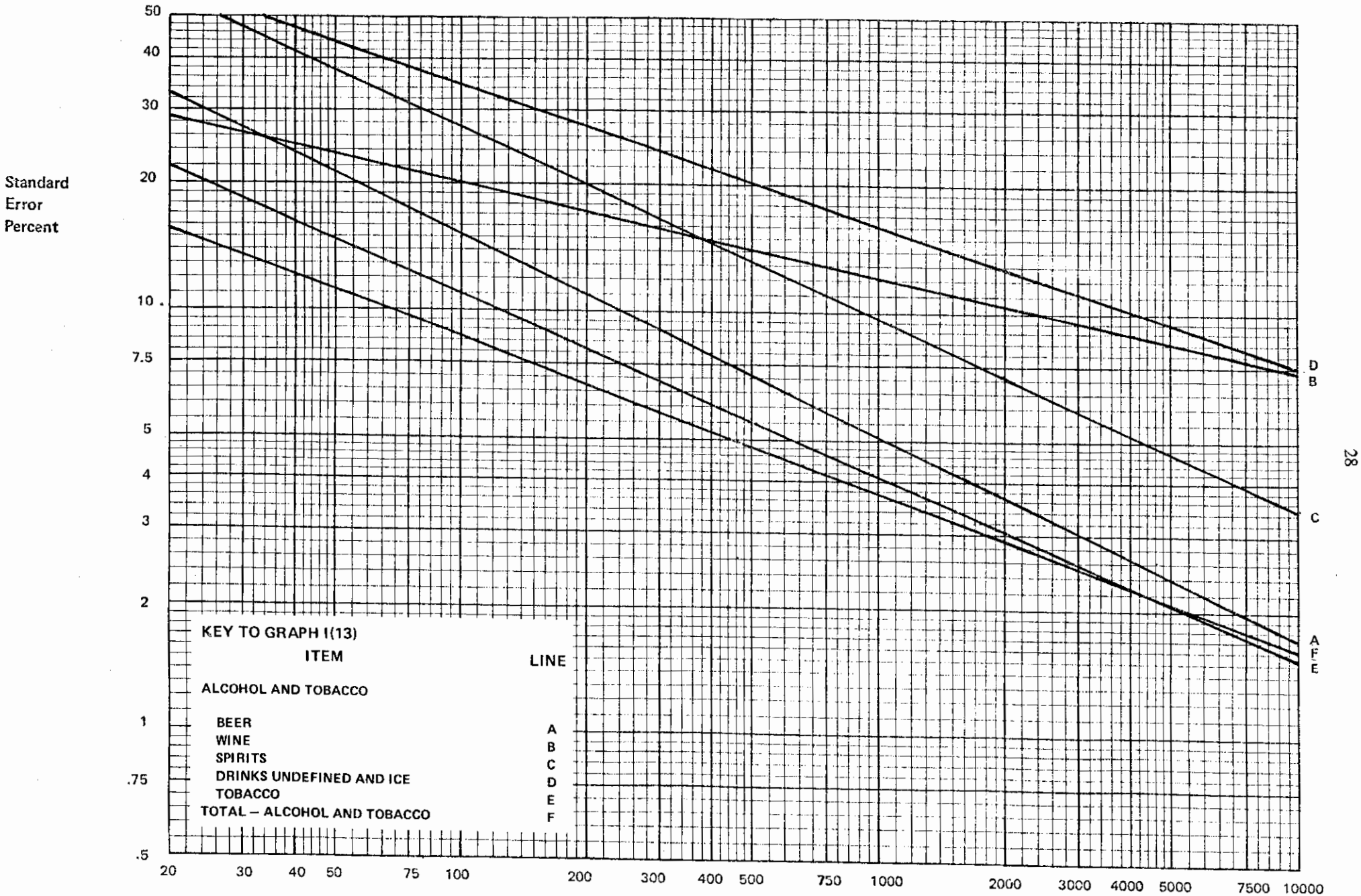
I(11) – EXPENDITURE ON FRUIT AND VEGETABLES



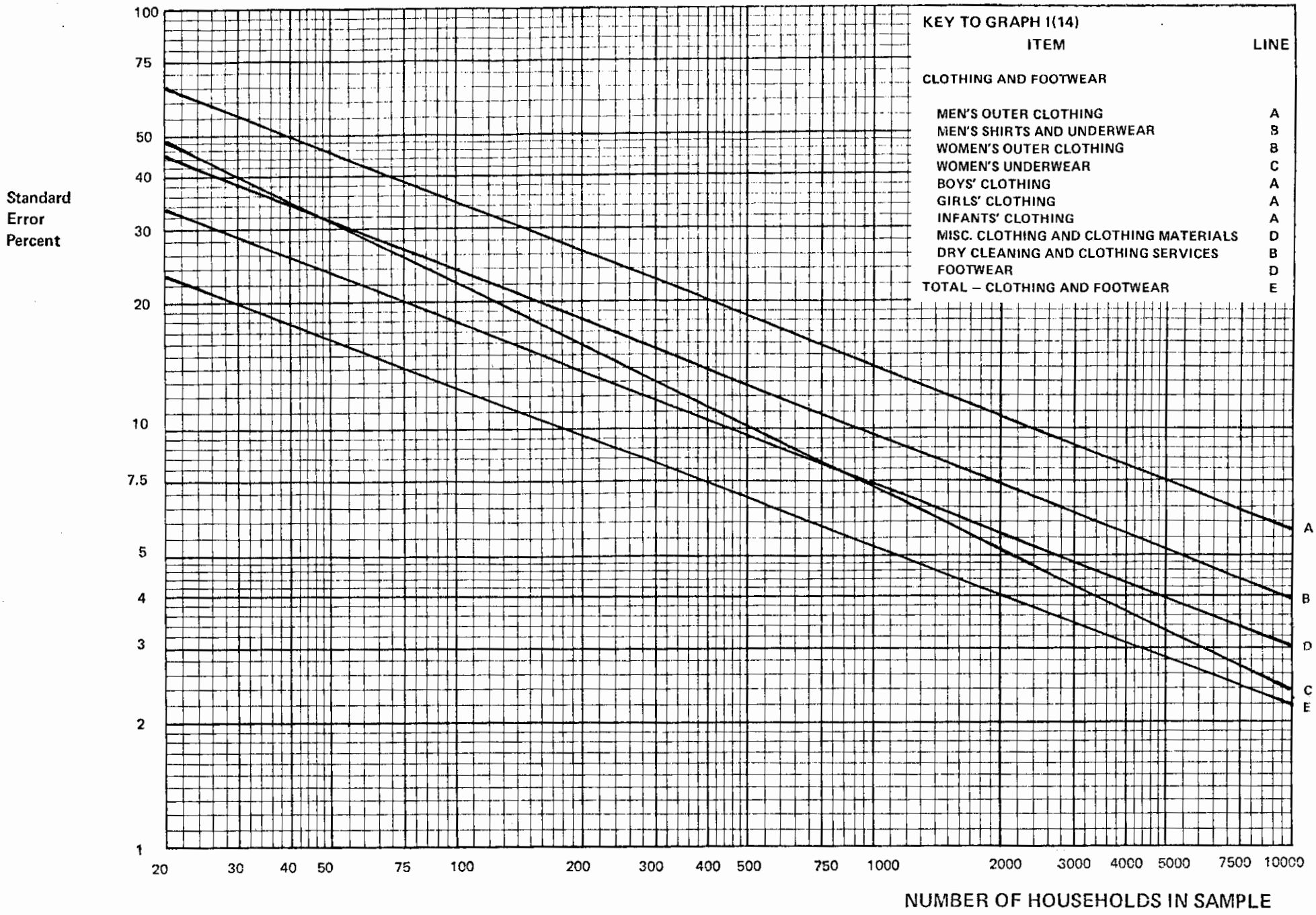
I(12) – EXPENDITURE ON OTHER FOOD



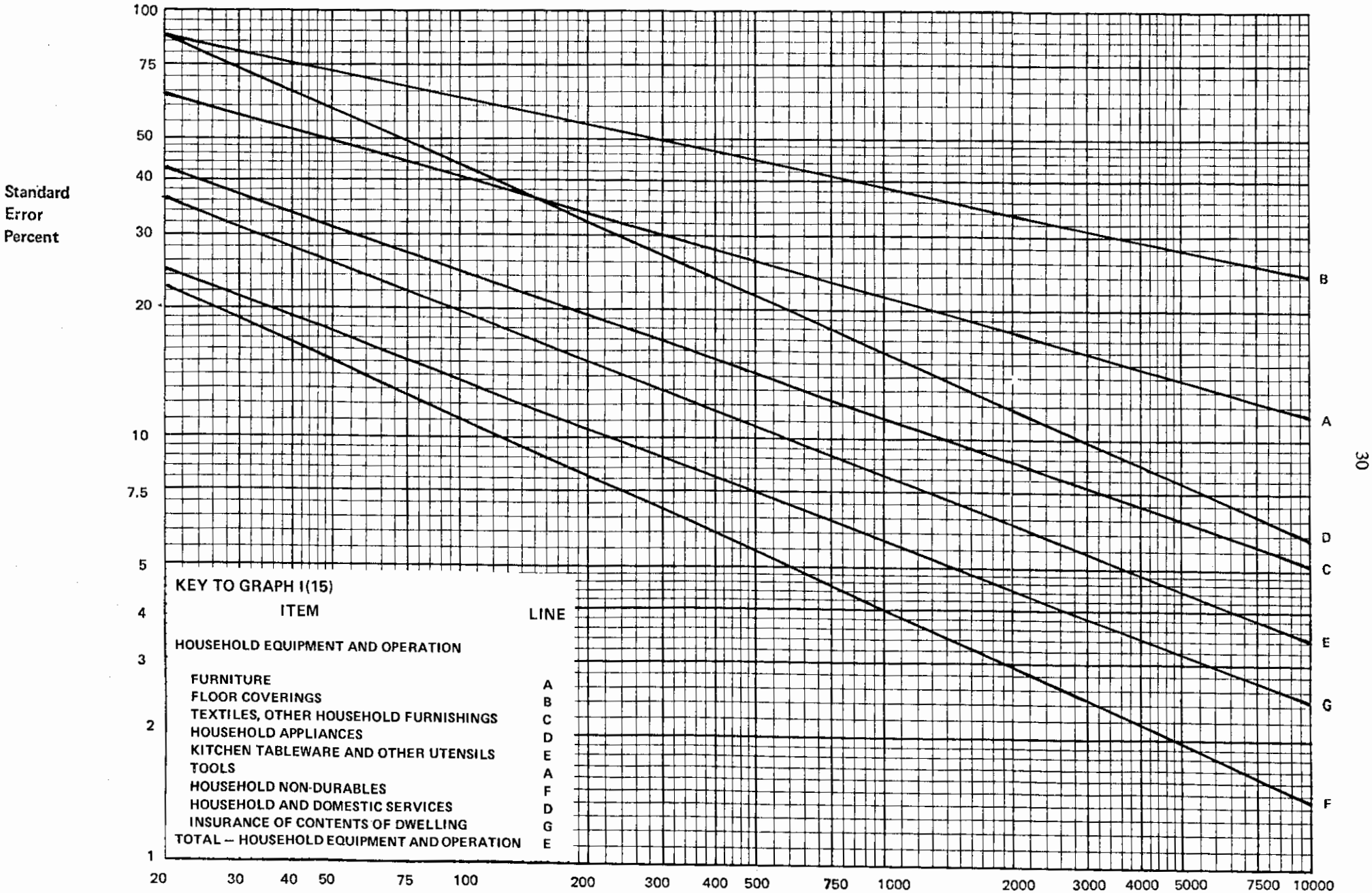
I(13) – EXPENDITURE ON ALCOHOL AND TOBACCO



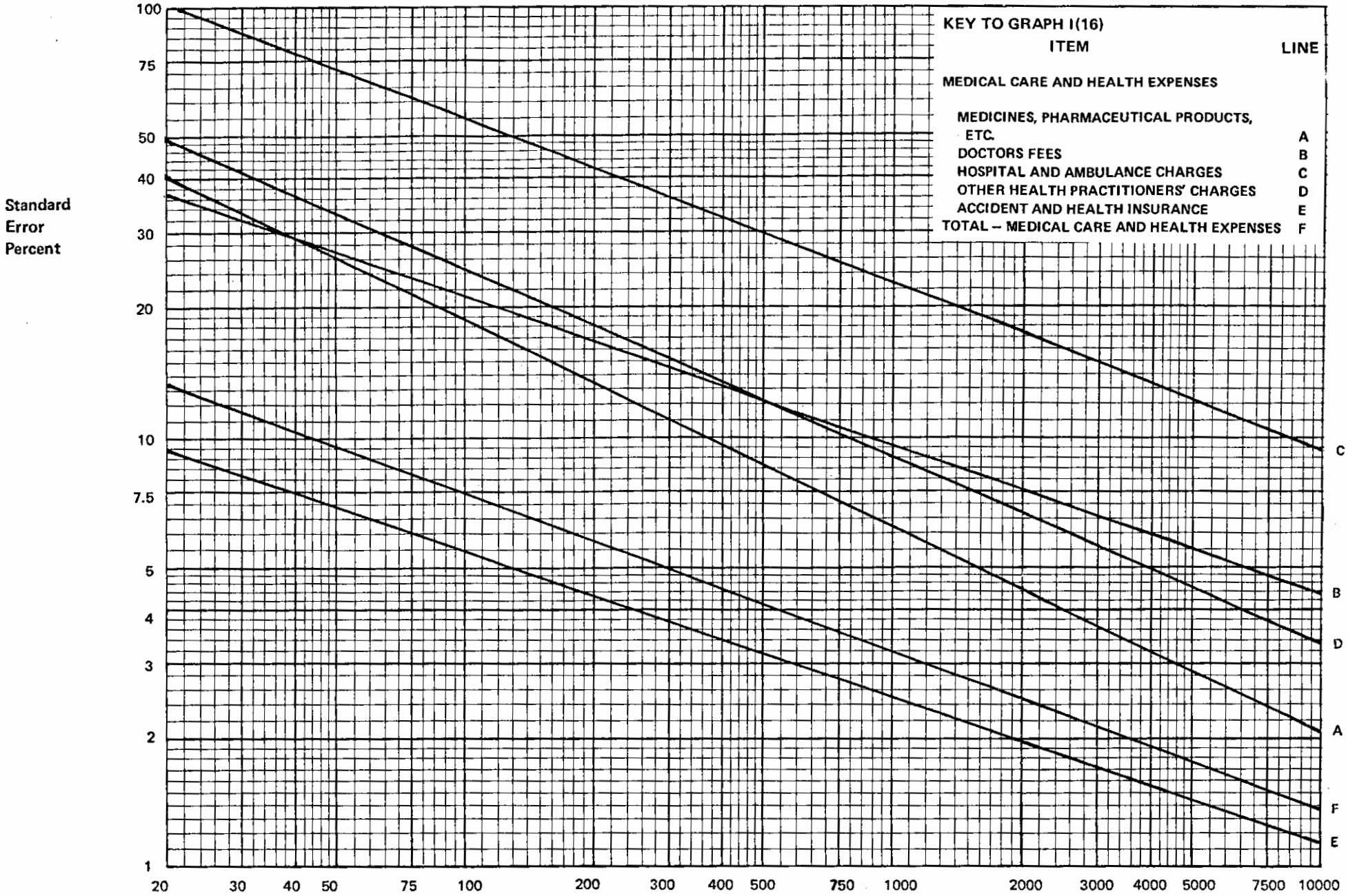
I(14) – EXPENDITURE ON CLOTHING AND FOOTWEAR



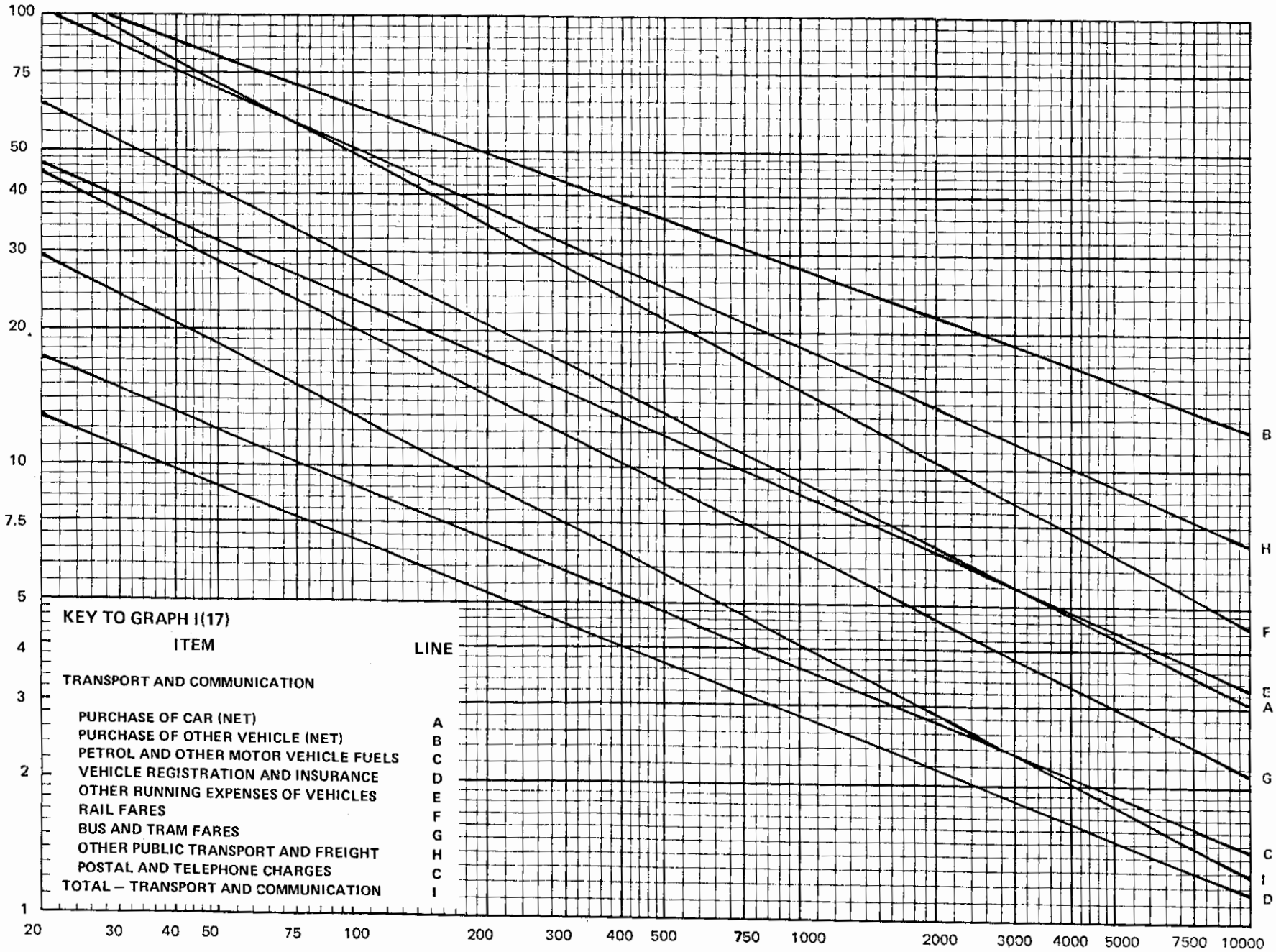
I(15) – EXPENDITURE ON HOUSEHOLD EQUIPMENT AND OPERATION



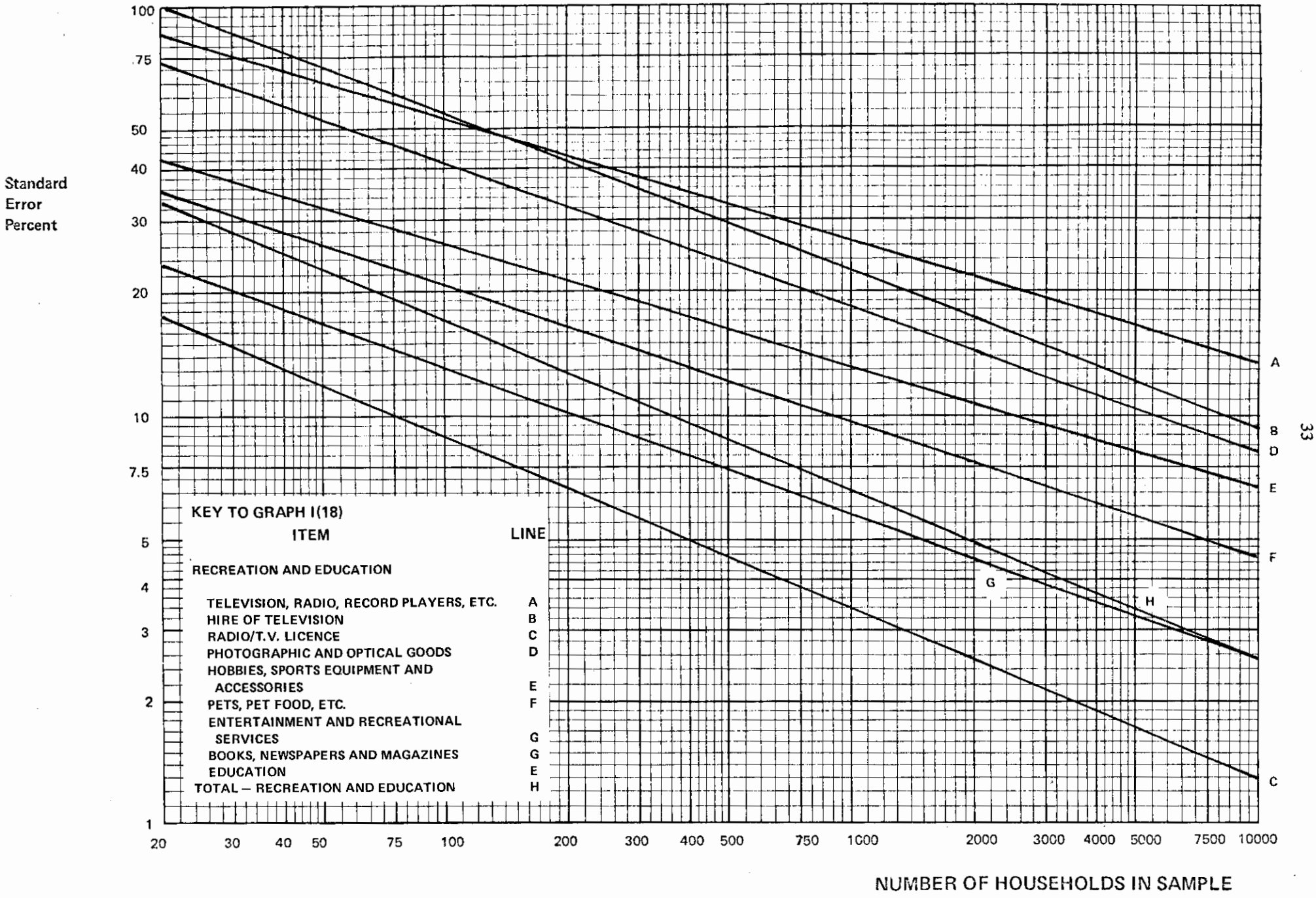
I(16) – MEDICAL CARE AND HEALTH EXPENSES



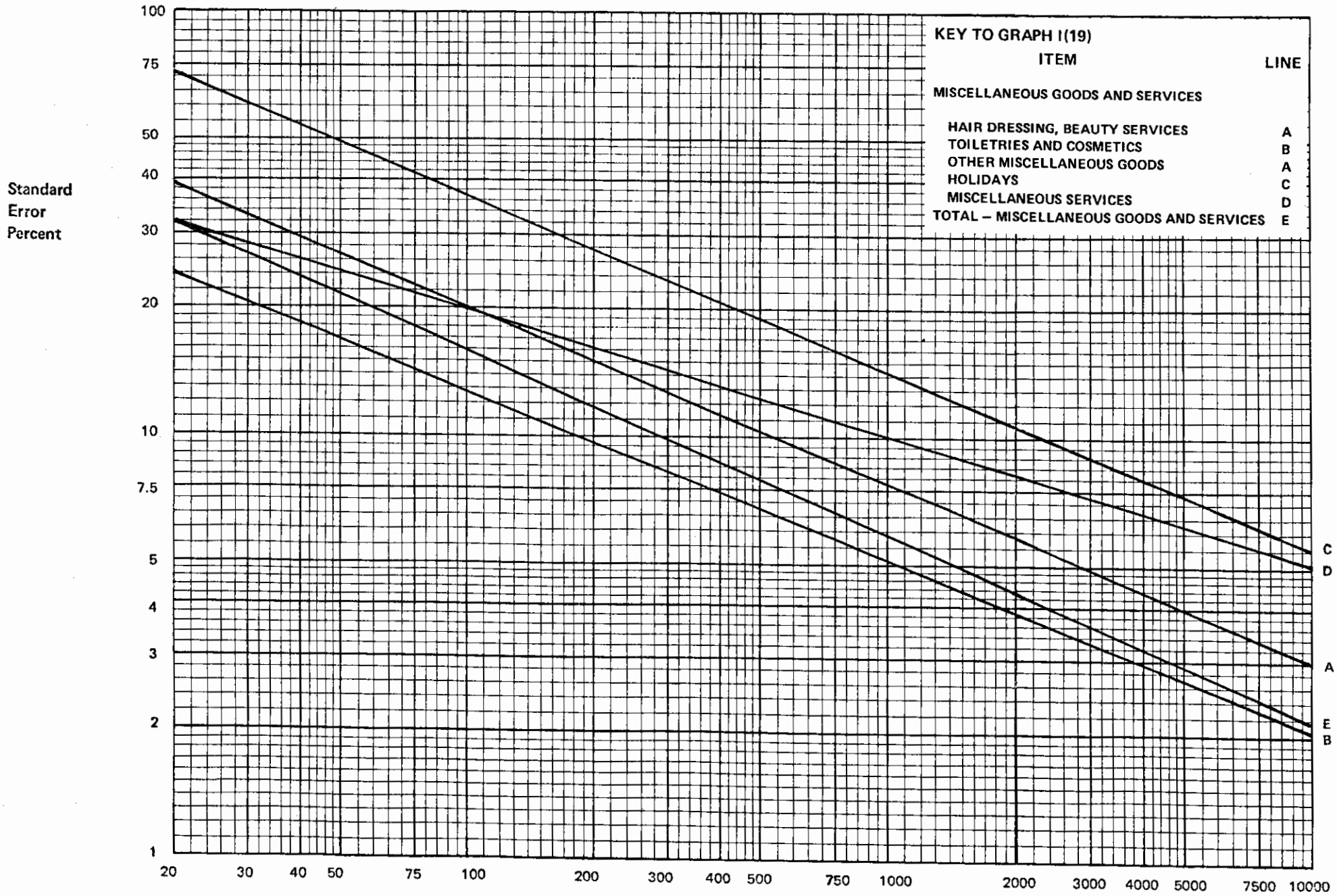
I(17) – EXPENDITURE ON TRANSPORT AND COMMUNICATION



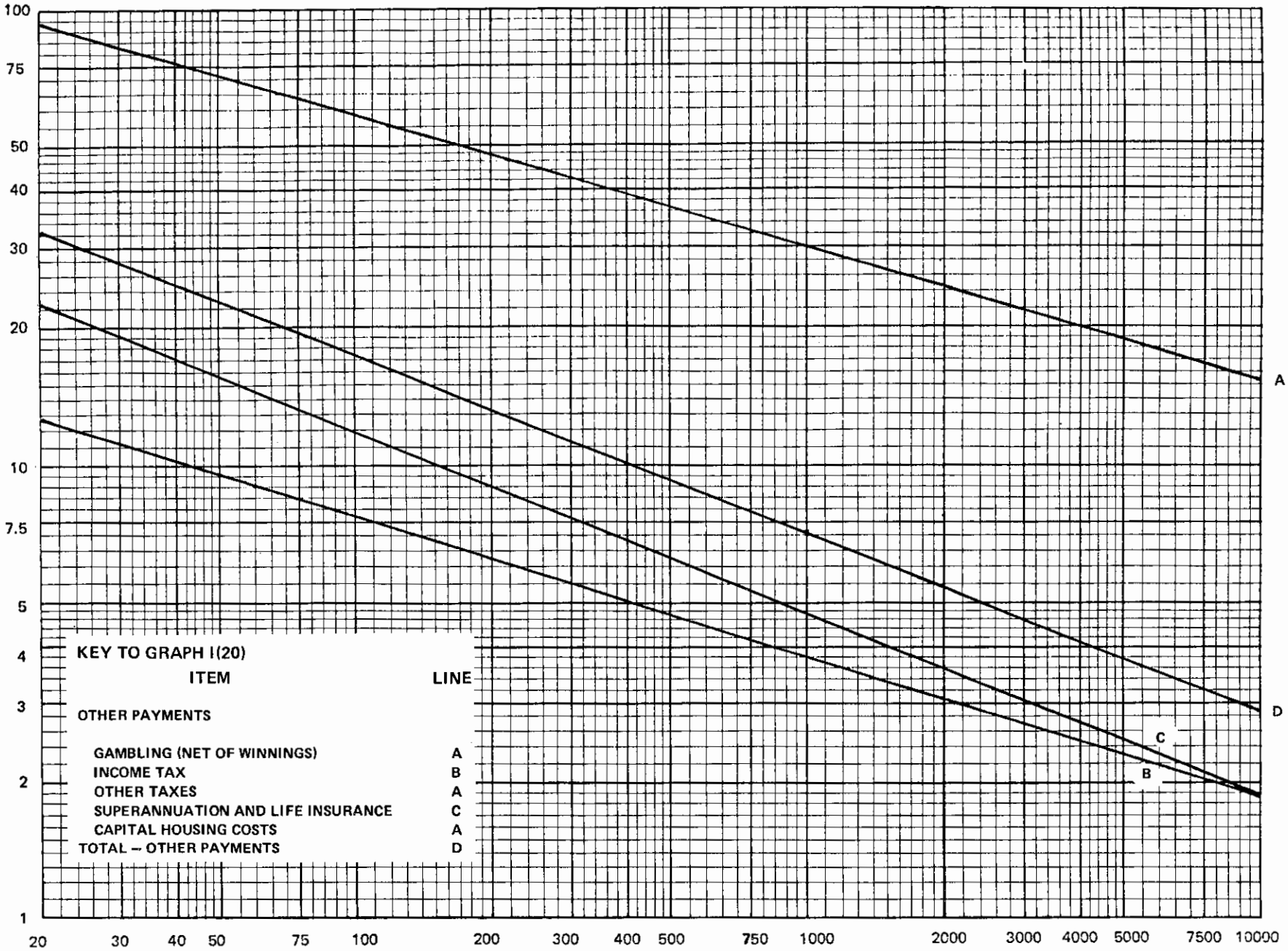
I(18) – EXPENDITURE ON RECREATION AND EDUCATION



I(19) – EXPENDITURE ON MISCELLANEOUS GOODS AND SERVICES



I(20) – OTHER PAYMENTS



2.2 GROUP II– NUMBERS OF HOUSEHOLDS AND NUMBERS OF PERSONS TABLES

This group contains Tables 7.1 to 7.8 (Numbers of Households and Cumulative Percentage of Households by Household Income), and Tables 7.27 to 7.31 (Numbers of Persons and Cumulative Percentage of Persons by Personal Income). Details of income were collected from all household members aged 15 years and over. For the purposes of the standard error graphs the term persons is taken to represent persons aged 15 years and over, income recipients, or employees (as defined in published tables).

For this group standard errors are presented in the form of graphs of standard error percent against the size of the estimate of number of households (for Tables 7.1 to 7.8) and number of persons (for Tables 7.27 to 7.31). The standard error percent for estimates of numbers can be found directly from the graph, while estimates of percentage must be converted to an estimated number for the determination of their standard errors.

The standard error percent for a particular estimate may be obtained as follows:

(i) Look up the estimate size (ie the actual figure published) in the table concerned.

(ii) Find the appropriate graph viz.

  (a) Estimates less than 50,000 – Graph II (1) on page 37.

  (b) Estimates greater than 50,000 – Graph II (2) on page 38.

(iii) Refer to the key on this graph to find the appropriate line.

(iv) Read the standard error percent from the graph.

To calculate the standard error percent in the case of a cumulative percentage it will first be necessary to convert the estimated percentage to an estimated number. This may be done as follows:

(i) Look up the total number which constitutes 100 percent of the group concerned. This may be found in the total row of the 'numbers' half of the table.

(ii) Calculate the estimated number using the formula

$$\text{Estimated Number} = \frac{\text{Total number} \times \text{Estimated percentage}}{100}$$

(iii) The standard error percent may then be found by reading the standard error percent corresponding to this number on Graph II (1) or II (2).

Examples

Estimated Numbers

Table 7.1 of Bulletin 7 shows that the number of households in Melbourne with weekly household income between $200 and $220 is estimated to be 58,800. Having determined from Table 3.1 of this bulletin that Group II is the appropriate group the standard error percent may be obtained as follows:

(i) The estimated number is 58,800.

(ii) The appropriate graph is Graph II (2) on page 38.

(iii) The appropriate line is line A (household line).

(iv) Looking up 58,800 on this graph gives the standard error percent as 5.6.

Thus the standard error is 3,300, the 67 percent confidence interval is 55,500 to 62,100 and the 95 percent confidence interval is 52,200 to 65,400.

Estimated Percentages

Table 7.27 of Bulletin 7 shows that the percentage of persons in Brisbane with weekly personal income less than $200 is estimated to be 93.1 percent. Having determined from Table 3.1 of this bulletin that Group II is the appropriate grouping, the standard error percent may be obtained as follows:

(i) The total row of Table 7.27 indicates that the estimated number of persons in Brisbane is 651,000, so the 93.1 percent with income under $200 constitutes 651,000 x 93.1/100 = 606,000 persons.

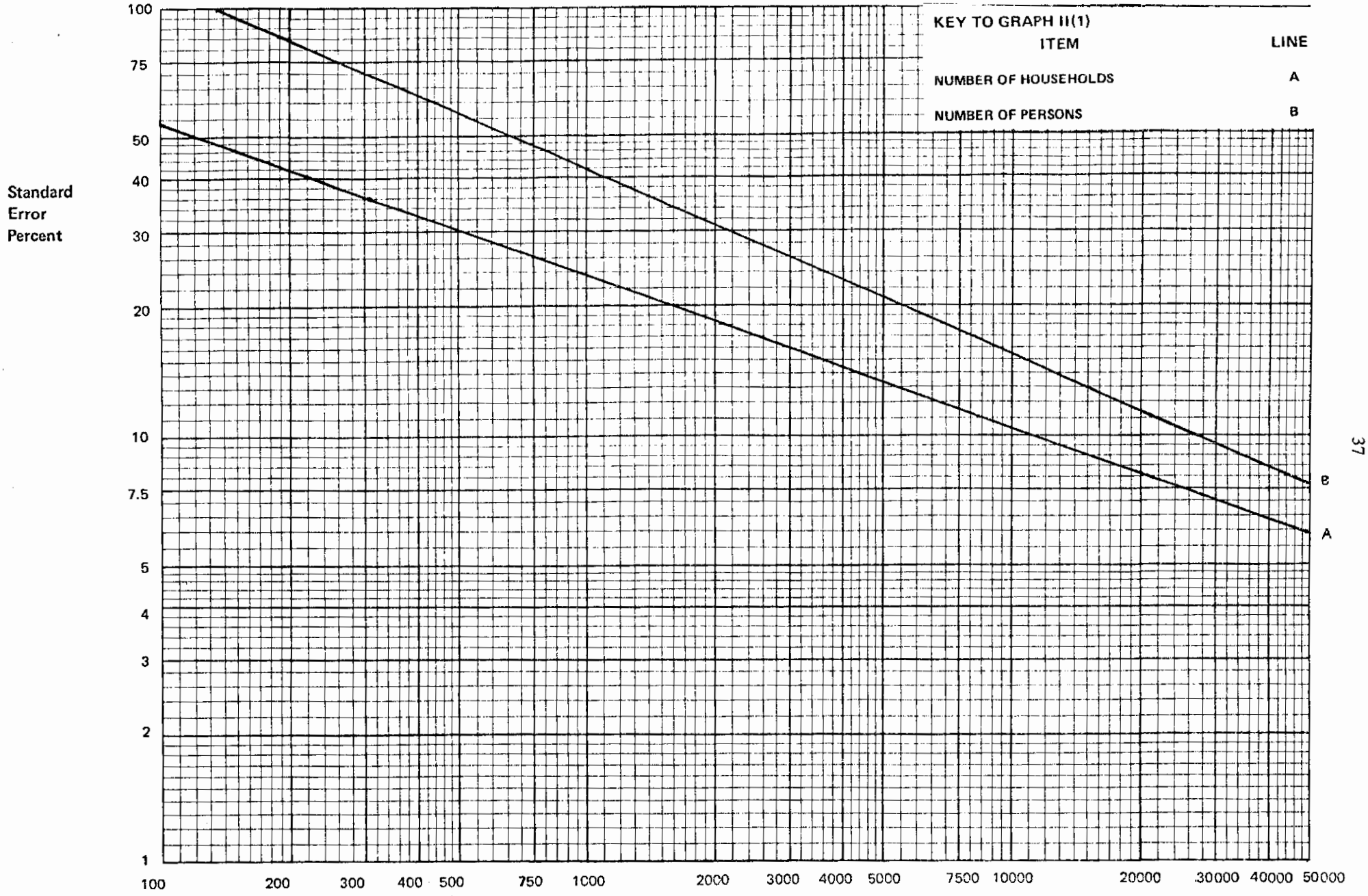
(ii) Graph II (2) on page 38 is the appropriate graph.

(iii) The appropriate line is Line B (persons line).

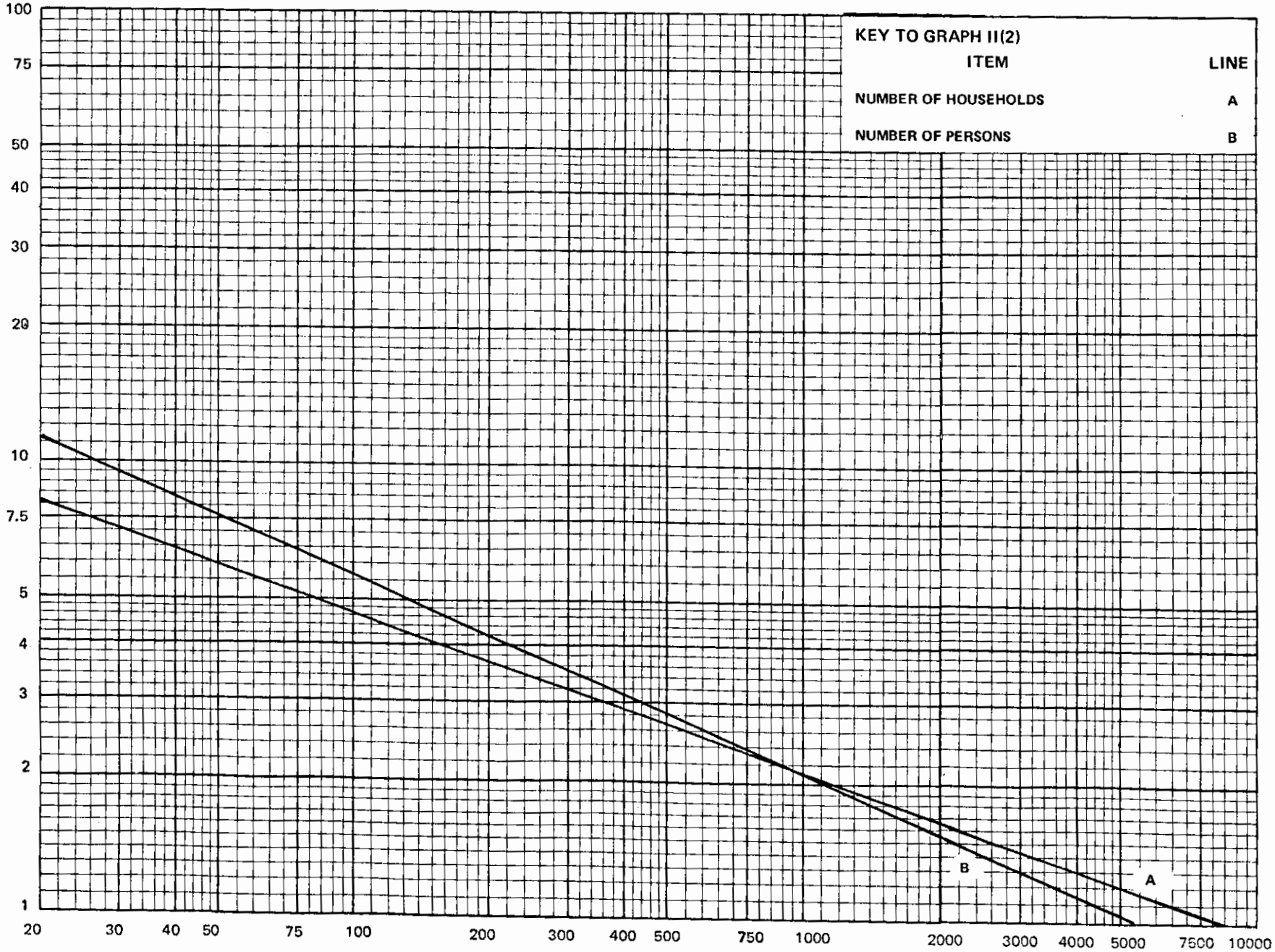
(iv) Looking up 606,000 on this graph gives the standard error percent as 2.6.

Thus the standard error of the percentage estimate is 2.6 x 93.1/100 = 2.4 percent. The 67 percent confidence interval is 90.7 percent to 95.5 percent and the 95 percent confidence interval is 88.3 percent to 97.9 percent.

II(1) – NUMBERS OF HOUSEHOLDS AND NUMBERS OF PERSONS – ESTIMATES LESS THAN 50,000



II(2) – NUMBERS OF HOUSEHOLDS AND NUMBERS OF PERSONS – ESTIMATES GREATER THAN 20,000



2.3 GROUP III – HOUSEHOLD AND PERSONAL INCOME BY SOURCE OF INCOME TABLES

This group contains Tables 7.10 to 7.17 (Household Income by Source of Income and Percentage of Total Household Income) and 7.32 to 7.37 (Personal Income by Source of Income and Percentage of Total Personal Income). Personal income may relate to persons aged 15 years and over, to income recipients or to employees. It should be noted that Table 7.9 although similar, was not included in this group as it is a combination of household and personal figures. Table 7.9 can be found in Group V.

For this group standard errors are presented in the form of graphs of standard error percent against the number of households in the sample on which the estimate is based for Tables 7.10 to 7.17, and the number of persons in sample for Tables 7.32 to 7.37. The standard error percent for the estimate of income or percentage of total income from a particular source may be found as follows:

(i) Find the sample size on which the estimate is based. This is shown on the top line of the table of Bulletin 7 in the same column as the estimate in question.

(ii) Find the appropriate graph, viz.

  (a) Household Income and Percentage of Household Income – Graph III (1) page 40.

  (b) Personal Income and Percentage of Personal Income – Graph III (2) page 41.

(iii) Look up the source of income (corresponding to average income or percentage of income) on the key to the graph to determine the alphabetic code of the line corresponding to the estimate in question.

(iv) Read the standard error percent from the graph using the sample size found in step (i).

Example

Table 7.32 of Bulletin 7 shows that the average weekly personal income from wages and salaries for all income recipients in Melbourne is estimated to be $80.94. Having determined from Table 3.1 of this bulletin that Group III is the appropriate group the standard error percent can be obtained as follows:

(i) The number of persons in sample is found to be 5,323 from the top line of Table 7.32.

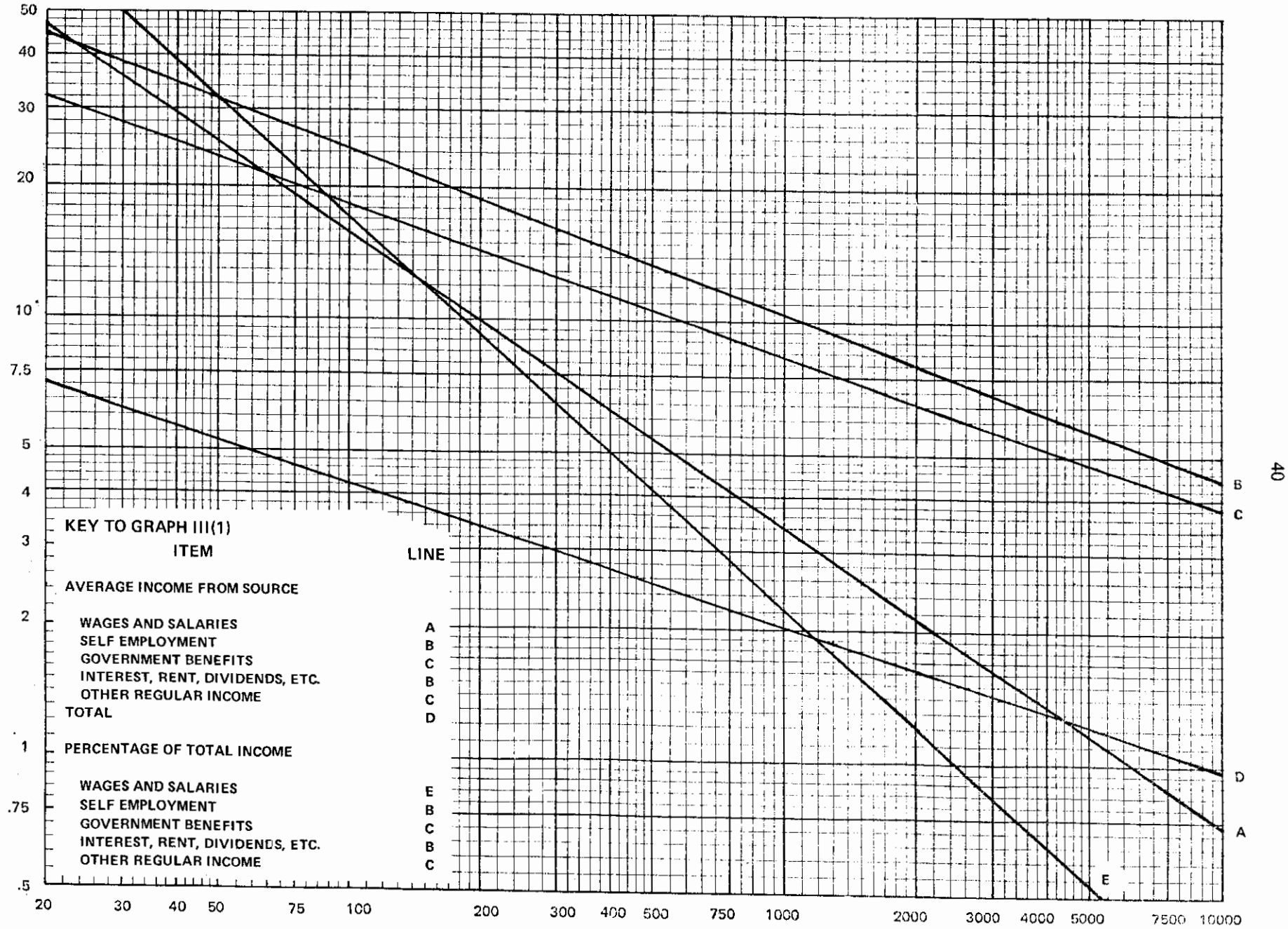
(ii) The appropriate graph is Graph III (2) on page 41.

(iii) The key on Graph III (2) indicates that Line A is the appropriate line.

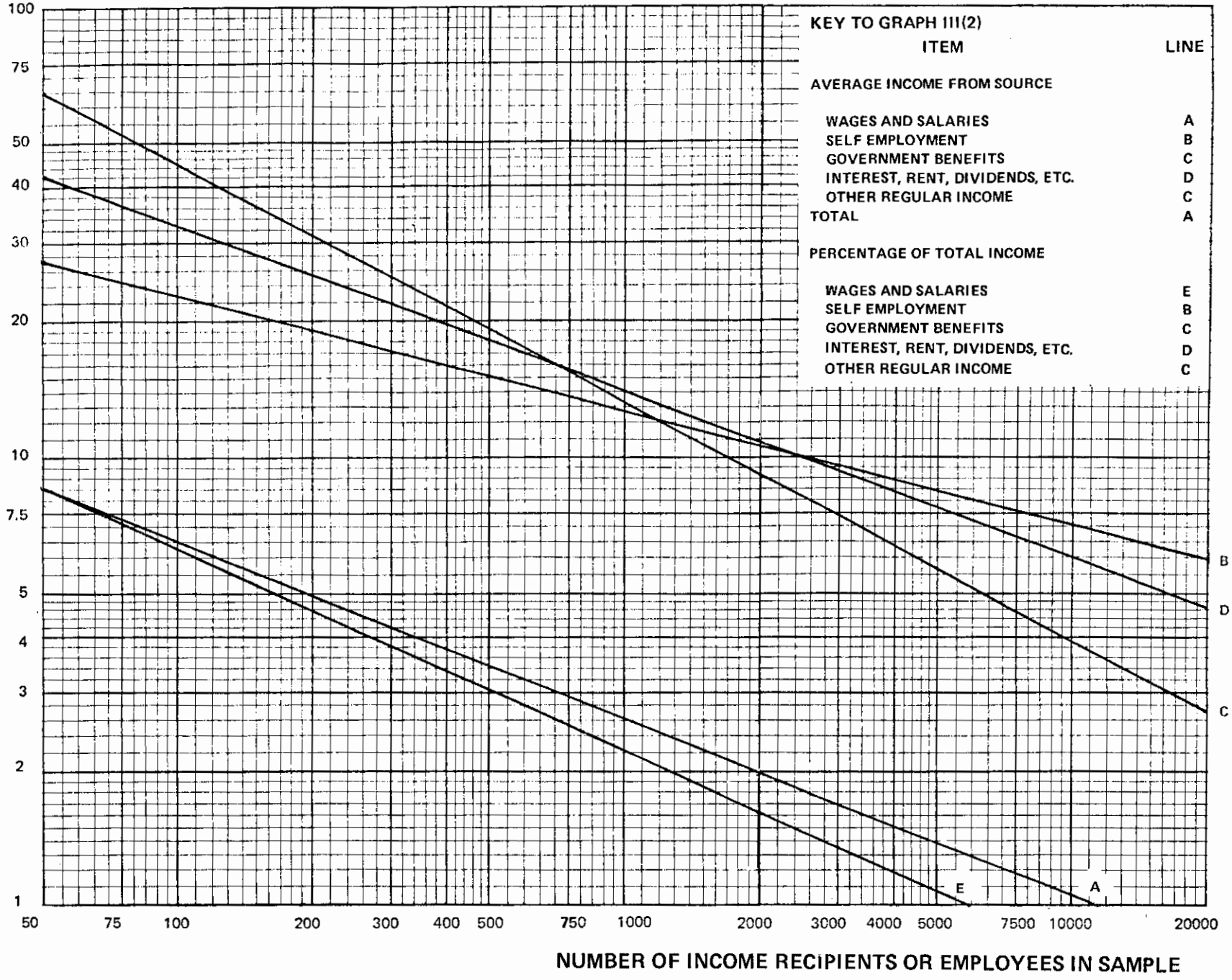
(iv) Looking up a sample size of 5,323 on this line gives the standard error percent as 1.4.

The standard error is therefore $1.13. The 67 percent confidence interval is $79.81 to $82.07 and the 95 percent confidence interval is $78.68 to $83.20.

III(1) – HOUSEHOLD INCOME FROM SOURCE AND PERCENTAGE OF TOTAL HOUSEHOLD INCOME



III(2) – PERSONAL INCOME FROM SOURCE AND PERCENTAGE OF TOTAL PERSONAL INCOME



2.4 GROUP IV – QUANTILES TABLES

This group contains Tables 7.18 to 7.26 (Quantiles of Household Income), Tables 7.38 to 7.43 (Quantiles of Personal Income), and Tables 8.17 to 8.22, 8.24 and 8.25 (Quantiles of Household Expenditure). Personal income may relate to persons aged 15 years and over, to income recipients or to employees. It should be noted that Table 8.23, although similar, was not included in this group, as there were insufficient points to draw accurate standard error graphs. Table 8.23 is included in Group V.

For this group standard errors are presented in the form of graphs of standard error percent against the number of households in sample for Tables 7.18 to 7.26, 8.17 to 8.22, 8.24 and 8.25, and number of persons in sample for Tables 7.38 to 7.43. The standard error percent for the estimate of a particular quantile may be found as follows:

(i) Find the sample size in the cell to which the quantile refers. This is shown in the left hand column of the table on the same line as the quantile in question.

(ii) Find the appropriate graph, viz.

  (a) Quantiles of Household Income – Graph IV (1), page 43.

  (b) Quantiles of Personal Income – Graph IV (2), page 44.

  (c) Quantiles of Household Expenditure – Graph IV (3), page 45.

(iii) Look up the quantile in question in the key on the graph to find the alphabetic code of the appropriate line.

(iv) Read the standard error percent from the graph using the sample size found in step (i).

Example

Table 7.18 of Bulletin 7 shows that the median household income in Perth is $174.51. Having determined from Table 3.1 of this bulletin that Group IV is the appropriate group, the standard error percent can be obtained as follows:

(i) The number of households in sample is found to be 1,089 from the left hand column of Table 7.18.

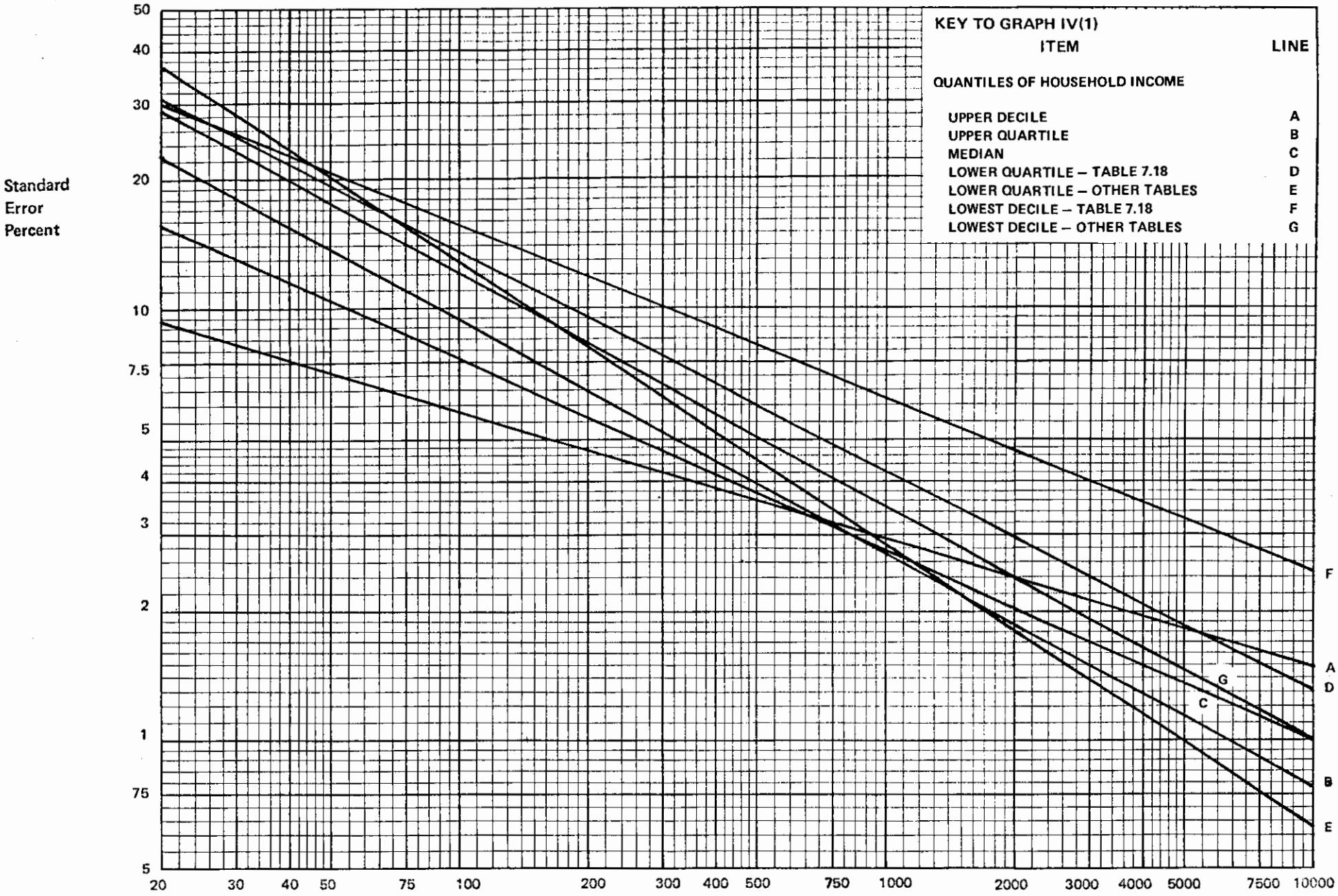
(ii) The appropriate graph is Graph IV (1) on page 43.

(iii) The key on Graph IV (1) indicates that Line C is the appropriate line.

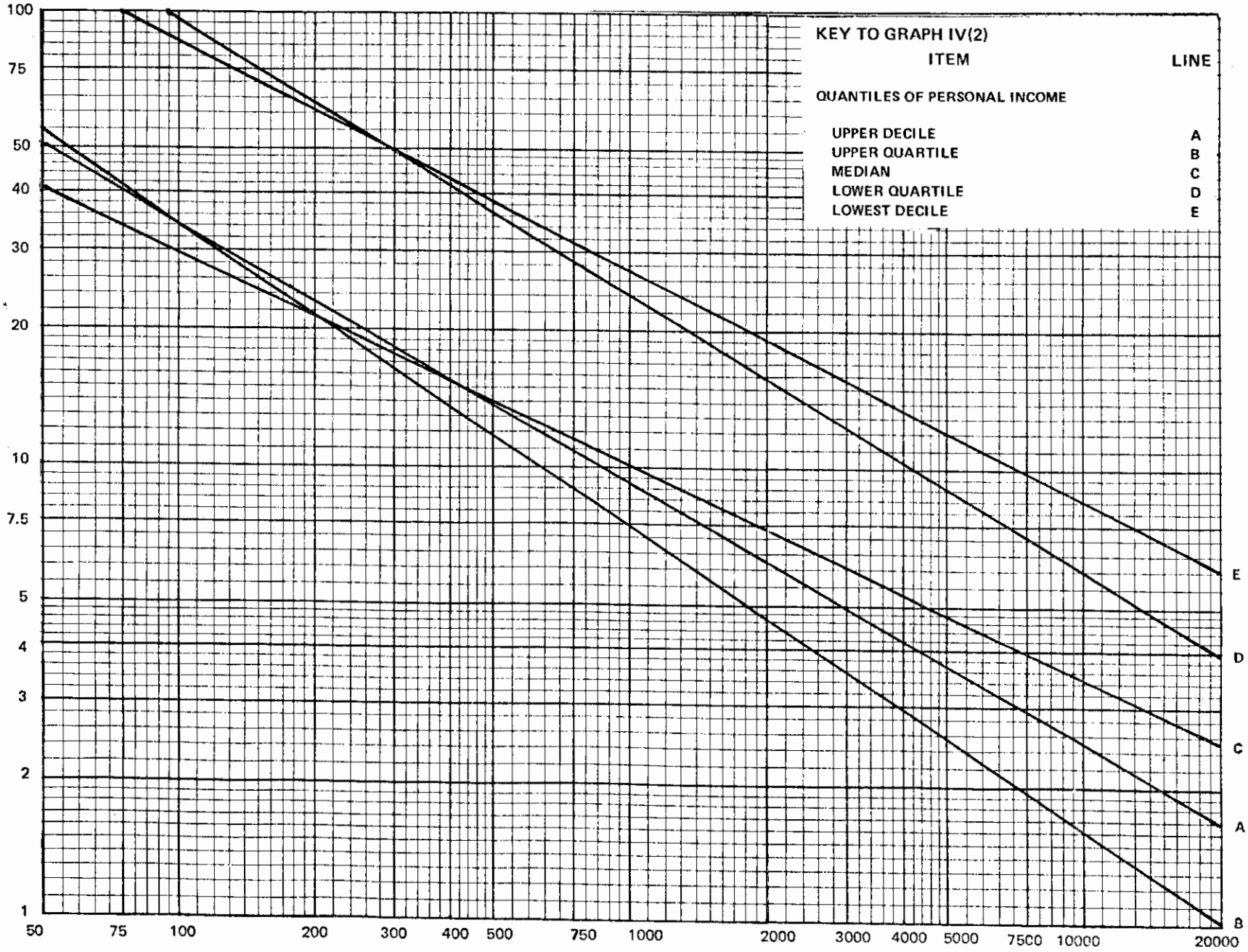
(iv) Looking up a sample size of 1,089 on this line gives the standard error percent of 2.7.

The standard error is therefore $4.71. The 67 percent confidence interval is $169.80 to $179.22 and the 95 percent confidence interval is $165.09 to $183.93.

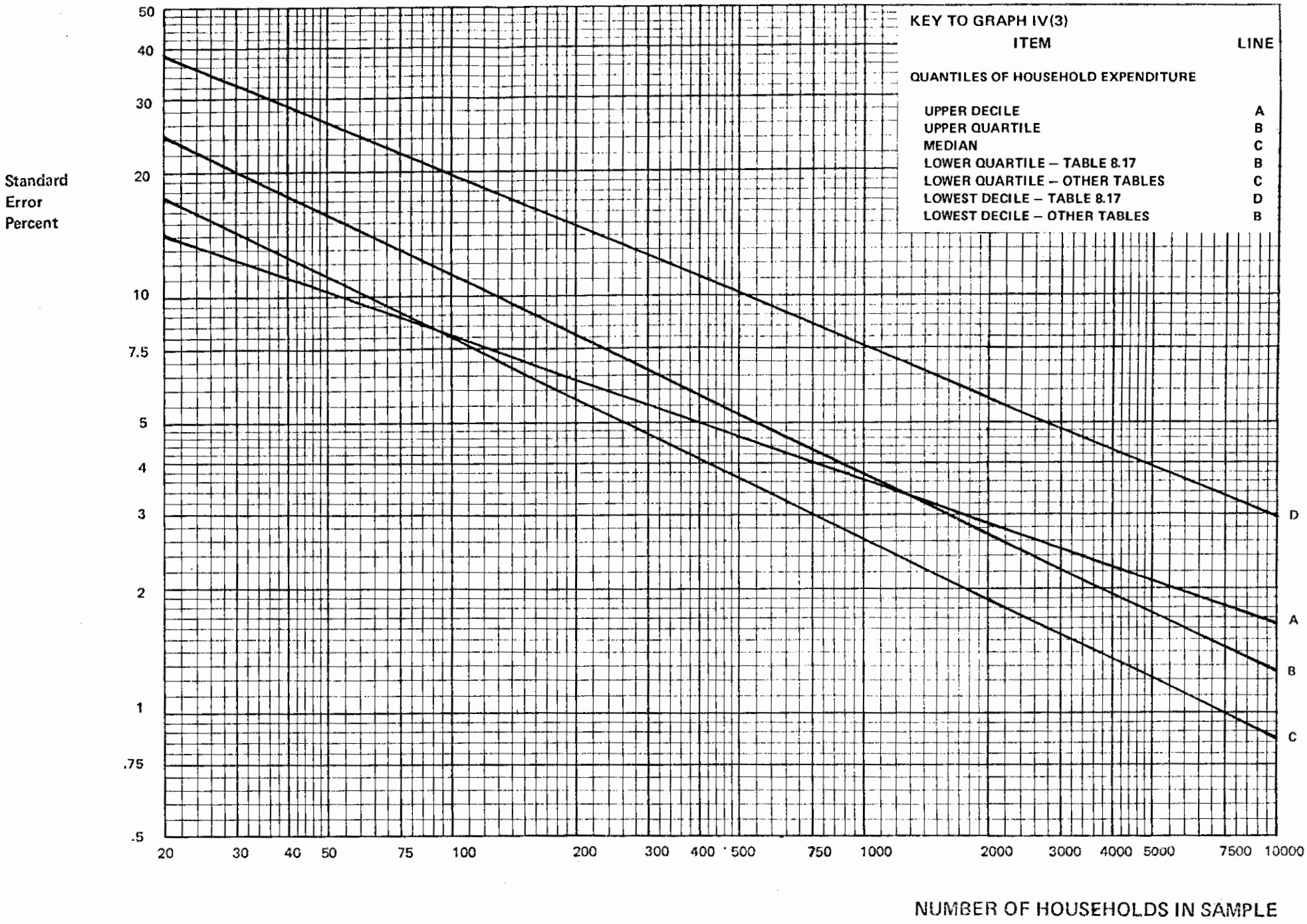
IV(1) – QUANTILES OF HOUSEHOLD INCOME



IV(2) – QUANTILES OF PERSONAL INCOME



IV(3) – QUANTILES OF HOUSEHOLD EXPENDITURE



2.5 GROUP V – TABLES OF STANDARD ERRORS.

This group contains the remaining two tables for which it was not possible to present standard errors in graphical form. The two tables are Table 7.9 – Average Income of Household Members and Percentage of Total Household Income by Source of Income and Table 8.23 – Quantiles of Household Expenditure by Household Expenditure.

The standard error percents for these tables are presented in tabular form below.

TABLE 3.3 STANDARD ERROR PERCENTS (FOR TABLE 7.9)

	Household Members				
Source of Income	*Head*	*Spouse of Head*	*Sons or Daughters of Head*	*Other Members*	*All Members*
	Average Weekly Income and Percentage of Total Household Income				
Wages and Salaries	1.2	2.0	3.7	6.0	1.0
Self Employment	5.0	10.1	17.9	28.1	5.2
Government Benefits	2.6	4.8	10.0	6.5	2.5
Interest, Rent, Dividends, etc.	4.7	11.4	18.6	21.6	4.4
Other Regular Income	4.7	4.1	14.1	13.4	3.7
TOTAL	**1.0**	**1.8**	**3.4**	**5.2**	**0.9**

NOTE: The standard error percents of Incomes and those of Percentages are approximately equal, and both may be found using this table.

TABLE 3.4 STANDARD ERROR PERCENTS (FOR TABLE 8.23)

	Quantile				
Broad expenditure item	*Lowest decile*	*Lower quartile*	*Median*	*Upper quartile*	*Highest decile*
Current housing costs	5.7	4.2	1.9	1.1	1.5
Fuel and power	2.7	1.1	0.9	1.1	1.3
Food	2.3	1.4	0.7	1.0	1.1
Bread, cakes and cereals	2.5	1.6	1.4	1.1	1.4
Meat and fish	4.1	1.9	1.5	1.3	1.6
Dairy products, oil and fats	2.7	1.7	2.3	1.2	1.4
Fruit and vegetables	3.4	1.7	2.1	1.4	1.4
Other food	3.1	2.4	1.5	1.5	2.2
Alcohol and tobacco	n.a.	7.4	2.6	1.8	2.2
Clothing and footwear	n.a.	5.6	2.7	2.3	2.3
Household equipment and operation	4.6	3.1	2.7	2.7	4.3
Medical care and health expenses	5.0	2.5	1.3	1.4	1.8
Transport and communication	5.5	2.5	2.1	2.3	2.4
Recreation and education	4.5	3.3	2.2	1.8	2.5
Miscellaneous goods and services	4.7	3.2	2.4	2.3	2.6
Total expenditure	3.6	1.5	1.0	1.1	1.5
Other payments	n.a.	5.5	2.0	1.9	0.7

n.a. not available

2.6 STANDARD ERRORS ON DERIVED STATISTICS – SUMS, DIFFERENCES AND RATIOS

The previous discussion has centred on the consideration of sampling errors associated with published survey statistics. Published figures may also be used to estimate other statistics such as the sum, difference, or ratio between two published survey estimates. For example, one may wish to consider the estimate of the difference between expenditure on 'Fuel and Power' in the March and June quarters. Such a figure is itself a survey estimate and is therefore also subject to sampling error.

These estimates are called derived estimates and their standard errors depend on the standard errors of the original estimates and on the correlation between the original estimates. The correlation depends on the nature of the component estimates and will therefore vary for different derived items. By assuming the correlation is zero, an approximate standard error can be calculated by using the following formulae:

(a) *Sums* – The approximate standard error percent on the sum $x_1 + x_2$ of two estimates x_1 and x_2 is given by

$$\text{SE\%}(x_1 + x_2) = \frac{\sqrt{(x_1 \times \text{SE\%}(x_1))^2 + (x_2 \times \text{SE\%}(x_2))^2}}{(x_1 + x_2)}$$

where SE% (x_1) and SE% (x_2) are the standard error percents on the estimates x_1 and x_2 respectively.

(b) *Differences* – The approximate standard error percent on a difference $x_1 - x_2$ is given by

$$\text{SE\%}(x_1 - x_2) = \frac{\sqrt{(x_1 \times \text{SE\%}(x_1))^2 + (x_2 \times \text{SE\%}(x_2))^2}}{(x_1 - x_2)}$$

(c) *Ratios*– The approximate standard error percent on a ratio x_1/x_2 is given by

$$\text{SE\%}(x_1/x_2) = \sqrt{(\text{SE\%}(x_1))^2 + (\text{SE\%}(x_2))^2}$$

Example

Tables 2.2 and 2.3 in Bulletin 2 show that average weekly household incomes in Sydney and Melbourne were \$208.14 and \$212.19 respectively, based on sample sizes of 2255 and 2544. The standard error graph for these figures shows that the standard error percents are 1.5 percent and 1.4 percent respectively. The estimate of the difference between the two estimates is \$4.05. Using the approximation above, the standard error percent of this difference is estimated to be

$$\text{SE\%} = \frac{\sqrt{(208.14 \times 1.5)^2 + (212.19 \times 1.4)^2}}{4.05} = 106.4\%$$

The standard error is therefore \$4.31.

Comments on Derived Statistics

(a) When the two estimates in question are based on independent sub–groups of the sample (eg different cities, quarters, income groups), the correlation will be near zero and the standard error approximations above will be good.

(b) When the two estimates in question are based on the same sample (eg in the same city, quarter, income group), but are estimates for different items (e.g. different expenditure items or income from different sources), care should be taken with the above standard error formulae. If the items are in a sense substitutes (eg 'Electricity' and 'Gas'), then the estimates will be negatively correlated, and the standard errors on sums will be overestimated but those on differences and ratios will be underestimated. If the items are likely to occur together or to have large values together (eg 'Vehicle Registration and Insurance' and 'Petrol and Other Motor Vehicle Fuels'), then the estimates will be positively correlated, and the standard errors on sums will be underestimated but those on differences and ratios will be overestimated.

(c) In some cases the purpose of calculating derived statistics is to obtain estimates of expenditure at a level of amalgamation in between a 'medium' item and a 'broad' item. For example expenditure on 'Alcohol' may be required rather than 'Alcohol and Tobacco' (broad) or the component medium items 'Beer', 'Wine', 'Spirits', and 'Drinks Undefined and Ice'. In such cases treating the estimates as a sum of medium items (eg 'Alcohol' = 'Beer' + 'Wine' + 'Spirits' + 'Drinks Undefined and Ice') and repetitively using the approximate formula for sums, will generally result in a more accurate standard error estimate than treating the estimate as the difference between the broad item and one or more medium items, (eg 'Alcohol' = 'Alcohol and Tobacco' – 'Tobacco').

APPENDIX 1 – NON – SAMPLING ERRORS

The following is a discussion of the non–sampling errors associated with the Household Expenditure Survey and the efforts made to reduce them and to measure their effects.

As outlined in the main text, the major sources of non–sampling errors are:

(a) errors due to non-response, ie because of differences in the characteristics and patterns of expenditure and income between respondent and non–respondent households.

(b) errors on the part of respondents and interviewers such as incorrect interpretation or wording of questions and inability or unwillingness to provide accurate information.

(c) errors in processing such as mistakes in the recording or coding of the data obtained.

(d) errors due to the collection of the data over an extended time period.

Each of these sources of error is discussed separately in the following.

Non – Response Errors

The effects of non – response errors can be minimised, firstly by taking measures to maximise response, and secondly by adjusting for non – response during the estimation phase. Considerable emphasis was placed on achieving high levels of response through intensive interviewer training, follow up calls by the interviewer during diary keeping and the use of incentives (eg diary keeping aids). The table below shows the sample size, the number of cooperating households and response rates for each of the seven cities.

RESPONSE RATES

City	*Sample size*	*Fully co-operating households*	*Response rate (%)*
Sydney	3445	2255	65
Melbourne	3467	2544	73
Brisbane	1584	1119	71
Adelaide	1360	983	72
Perth	1315	1089	83
Hobart	731	562	77
Canberra	690	543	79
All Capital Cities	**12592**	**9095**	**72**

Considering the nature of the survey these response rates were considered satisfactory. Surveys conducted overseas have achieved similar response rates.

Section 6.3 in Bulletin 1 describes the method used to adjust for non–response using the technique known as 'post–stratification'. The procedure consists of assigning each responding and non-responding household to one of three dwelling type and household size 'post–strata' within each stratum and quarter, and adjusting the weights of responding households within a post–stratum according to the number of non–respondents in that post–stratum. The procedure effectively imputes for non–respondents on the basis of respondents of the same dwelling type and household size. For a detailed specification of the calculation of weights see Appendix 4.

The non–response adjustment is most efficient when:

(a) There are large differences between post–stratum response rates.

(b) The post–strata are different with respect to the characteristics being measured.

The following table gives an example of the differences between post–stratum response rates and levels of average weekly household expenditure (based on the September quarter's sample in Sydney).

POST–STRATIFICATION

Post–stratum	*Response rate*	*Average weekly household expenditure*
Houses with 1 or 2 persons	64%	$92.89
Houses with 3 or more persons	57%	$187.54
Flats	79%	$144.01

The figures suggest that the amount of non–response bias removed may be quite large, confirming the results of analyses based on earlier pilot tests. However the effect of post–stratification on the non–response error is only to remove that component of the non–response bias due to differences between post–strata. The bias remaining is that due to differences between respondents and non–respondents within post–strata. Without follow–up of non–respondents, it is not possible to provide an indication of the size of this bias, but it is believed to be small relative to the standard errors on estimates.

Response Errors

Response errors have three main sources:

(a) Deficiencies in the wording of questions (including inconsistent or inadequate definitions), or in questionnaire design and interviewer instructions.

(b) Deficiencies in interviewing technique (eg deviating from question wording, failing to follow instructions or failing to understand the reply, or making a mistake in recording the answer).

(c) Inaccurate reporting by the respondent because of misunderstanding of the question, a poor record–keeping system, inability to recollect the required data or deliberate incorrect answering.

In the design and conduct of the survey, considerable effort was made to minimise such errors by careful questionnaire design, extensive field testing, intensive interviewer training coupled with strict quality controls on their work, and by encouraging respondents to refer to their records whenever possible. In addition considerable pilot testing was carried out to determine the most efficient survey methodology from the point of view of minimising response problems. For example, investigations were made of the effects of diary and recall collection methods, of different types of interviewers and of different designs of the diary.

Some attempts were made to measure the response errors in the Household Expenditure Survey. These included a reinterview programme based on a subsample of the survey respondents, and other analyses of the effects of particular aspects of the collection procedures. These are discussed below.

Reinterview Programme

The reinterview programme was set up to assist in revealing areas of the questionnaires and survey procedures which needed to be revised for future surveys. It was also hoped that the reinterview programme would be able to measure the frequency and size of response errors. The programme consisted of reinterviewing one household from each interviewer's workload and comparing the answers given at the reinterview to those given at the original interview. Where differences occurred, an attempt was made to reconcile the differences.

Overall the reinterview programme provided the following estimates of response errors (assuming the data obtained from the reinterview programme is correct). These figures are percentages of the total number of questions answered in the reinterview programme.

Original recorded answer correct	: 93.3 percent
No answer originally recorded	: 3.5 percent
Original recorded answer incorrect	: 3.2 percent

Thus, in general, response errors appeared to be quite small although their size varied from question to question. When questions were considered separately and tested for bias, very few were found to have significant bias.

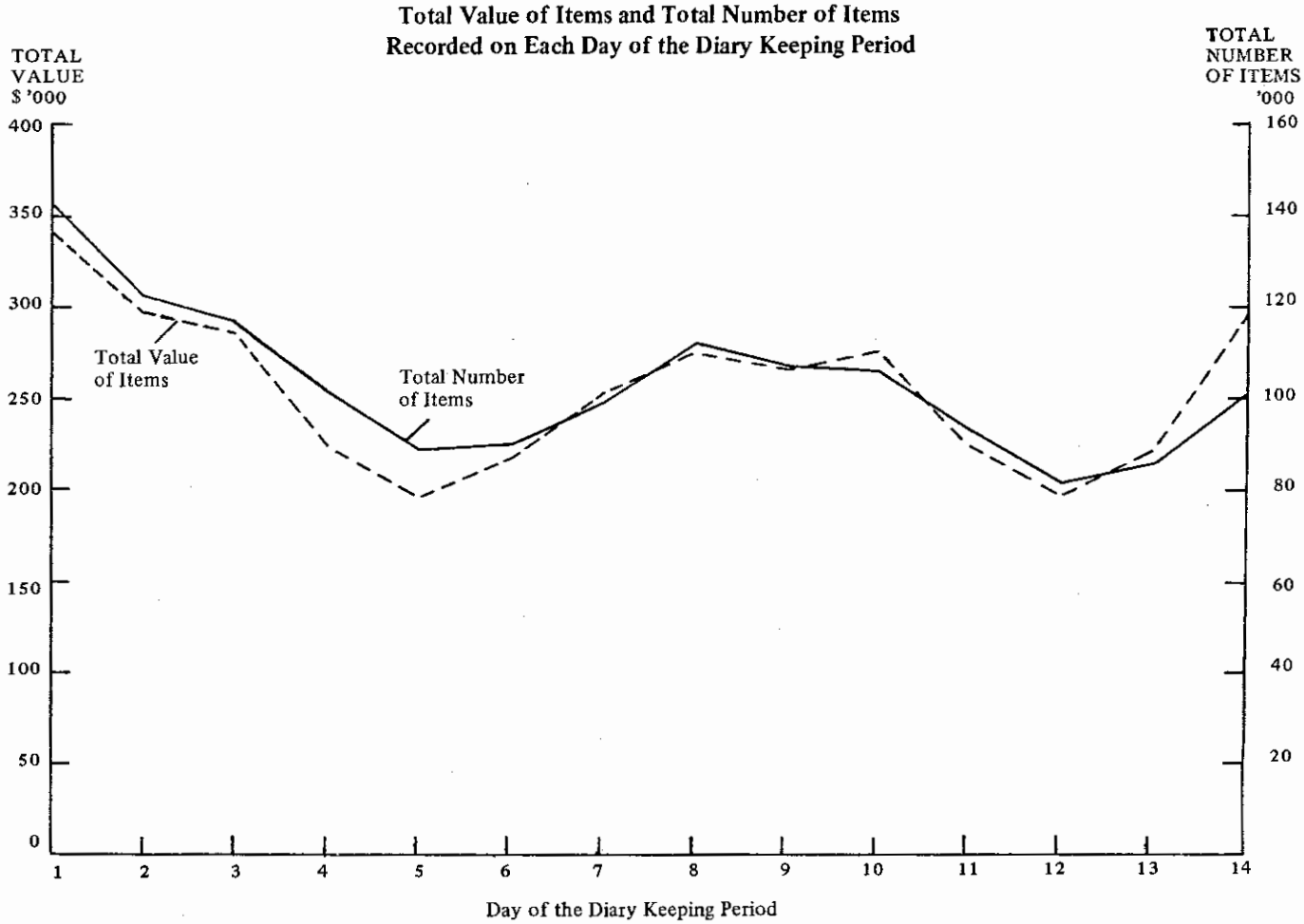
Other Quality of Data Studies

A number of studies are being undertaken to assess the quality of the data collected by the survey. While not all studies have been concluded, the results of some are available and are summarised below.

(a) Pattern of diary expenditure recording

A study was undertaken to ascertain if a pattern was evident in the incidence of recording payments over the diary keeping period. Data on the number of items recorded and on the value of items recorded in each day of the diary were examined. The following graph of the total number of items and the total value of expenditure recorded on each of the 14 days of the diary keeping period shows that a distinct pattern did exist. The peaks in the pattern coincide with the times when the interviewer was in contact with the household.

Total Value of Items and Total Number of Items Recorded on Each Day of the Diary Keeping Period



In an attempt to explain this pattern an allowance was made for the effects of trading days. Here the actual day of the week (ie Monday through Sunday) was given an appropriate weight and the study was repeated. While this did smooth out the reporting fluctuations to some extent, the systematic pattern still existed.

The existence of a distinct pattern in the reporting of expenditure in the diary may have implications on the accuracy of the estimates. It is not clear however, whether the peaks measure the 'true' level of expenditure (and thus the troughs represent an underestimate) or vice versa, or if the 'true' level of expenditure is somewhere between these extremes.

(b) Expenditure on alcohol

As pointed out later, the Household Expenditure Survey estimate of expenditure on alcohol understates that presented in the National Accounts. Although many possible explanations for this apparent understatement exist (including inability to recall expenditure, and the 'social undesirability' of alcohol), these are difficult to quantify. However, a component of the apparent understatement that can be quantified relates to meals eaten out in restaurants, hotels, clubs etc.

Where alcohol was consumed as part of a meal, and expenditure on alcohol was separately recorded then the payment for alcohol was coded to the relevant alcohol item. However, if the alcohol component of the meal was not separately identified the whole payment was coded to 'Meals in restaurants, hotels, clubs etc' (commodity code 248*). Thus a proportion of commodity code 248 will relate to alcohol payments. A study was undertaken in an attempt to estimate this component. Approximately 2.4% of the household diaries were examined and a subjective split of alcohol was made for households that did not separately identify alcohol in meals eaten out (this was assisted by type of outlet, and the characteristics of households that did separate alcohol and food components of meals). The results indicated that if the alcohol payments included in meals eaten out were included in thier appropriate alcohol grouping, then average weekly household expenditure on alcohol would have increased by approximately 4 percent.

A number of other studies are planned, to examine the quality of the data and to assist in the design of the next survey. These include: the optimal length of the diary keeping period; a comparison of diary and recall responses for those items collected by both means; and the use of the tick-box methodology to collect income data.

Processing Errors

Processing errors may occur anywhere between the supplying of information by respondents and final compilation of statistics into tables. Processing errors may occur at the following points in the processing system:

(a) miscoding of items by office processors.

(b) failure of editing at various stages to detect and correct errors in the data which could reasonably have been corrected.

(c) errors occurring in the transfer of the data from the original forms to magnetic tape.

(d) errors during the computer processing stage.

A number of steps were taken to minimise these types of errors at various stages of processing:

(i) *Edits :* Extensive edit checks were made at different stages in processing to detect processing errors, as well as response errors. These edits were carried out at three different levels:

(a) Interviewers checked recorded data using a form designed for this purpose before the final contact with the household was made. This enabled them to clarify apparent inaccuracies and inconsistences with respondents before the forms were passed on to the central processing centre. This was important as it was not possible to re-contact respondents once the forms had been sent to the central processing centre, as names and addresses had been removed by this stage.

(b) Specially trained processors in the central processing centre checked the data for completeness, consistency and unusual, duplicated or possibly inaccurate entries.

(c) Computer editing was carried out involving checking the logic of sequences followed in the forms, that necessary items were present, and that specific values lay between certain ranges. The main purpose of these edits was to detect keystroke errors that may have occurred when the data was transcribed onto magnetic tape and any incorrect relationships between items of data that may have been missed during clerical editing.

(ii) *Processing Quality Control:* A check was maintained on the quality of office processing by checking a sample (approximately 5 percent) of coded and edited household records. The main purpose of these checks was to monitor the performance of the processors to allow frequently occurring errors to be detected and corrected.

(iii) *Data File Checks:* Tabulations were obtained from the data file showing the distribution of households for different characteristics and the distributions of the values of some data items. These were used as a final check on the content of the data file and to identify unusual values which may have significantly affected estimates. If these unusual values were found to be in error then the data file was amended.

* See Appendix I Bulletin 1 (Catalogue No. 6507.0 Previously Ref. No. 17.19).

(iv) *Frequency of Coding Errors:* As the majority of clerical coding was of items recorded in the diary, a small experiment was conducted to measure the significance of coding errors in the diary. All items in the diary were coded to one of 300 fine level expenditure items. The test independently recoded a sample of the diaries (consisting of 0.5 percent of households and approximately 10,000 diary items) and then resolved any discrepancies between the recodes and the original codes. The test indicated that approximately 1.9 percent of items were wrongly coded at the 300 item classification level. Because data was reaggregated in the publication tables this figure reduced to 0.5 percent at the broad level of expenditure and 1.1 percent at the medium level.

Extended Collection Period

The methodology of the survey entailed the collection of data over a 12 month time period. If all the household variables collected in the survey remained constant during the time reference period then the length of the collection period would not have any implications for the survey estimates. Because the survey was conducted over an inflationary period, prices and incomes did change significantly during the collection period. These price change effects are thought to have been most significant in the classification of households. For example, two households with similar characteristics, one selected towards the beginning of the survey period and the other towards the end may have been classified into different income groups solely because of the changes in wage levels between the two collection dates.

A study was undertaken to examine the effect of the extended collection period on the estimated pattern of expenditure. The method used entailed the adjustment of each payment or receipt by a household according to prices prevailing in the December quarter 1974. One exception to this was the treatment of income from self-employment where the information was adjusted to the 1974–75 financial year. In addition, it was assumed that payments made for items collected on recall (eg medical expenses in respect of the previous 3 months, purchase of motor vehicles in respect of the previous 12 months) were made close to the time of interview.

The adjustments to individual data items were made on the basis of movements in a number of indicators. These were:

. Components of Consumer Price Index (for adjusting expenditure).
. Average weekly earnings per employed male unit (for adjusting wages and salaries).
. Components of National Accounts (for adjusting business income).
. Pension rates (for adjusting income from Government Benefits).

As price movements may have varied across Capital Cities, different movements were used for different cities wherever possible.

The table on the following page shows the original and adjusted patterns of expenditure for the broad expenditure groups (expressed in terms of percentage of total expenditure) for each of the six income ranges and for all households. Although the original and adjusted patterns differ within the income ranges, they are very similar for all households combined. This tends to support the hypothesis that the most significant effect of price changes is in the way households were classified. This is not necessarily restricted to the income variable.

Comparison of Household Expenditure Survey With Data From Other sources

Because the Household Expenditure Survey was the first comprehensive survey of its kind to be conducted in Australia, the amount of data with which valid comparisons could be made was limited. It was found that, although data on expenditure and income of the private sector existed, this data had differences which were difficult or impossible to allow for. These included factors such as the data being collected over a different period or being based on a different population or data unit. The collection period was a particular problem as price changes were significant during the period of the survey.

In comparing the survey data with data from other sources, allowance was made as far as possible for differences in scope, coverage, definition etc. This meant however, that generally comparisons could only be made at a broad level.

PERCENTAGE OF TOTAL EXPENDITURE (PUBLISHED AND ADJUSTED ESTIMATES) – ALL CAPITAL CITIES

Commodity or service	*Weekly household income*													
	Under $80		*$80 and under $140*		*$140 and under $200*		*$200 and under $260*		*$260 and under $340*		*$340 or more*		*All households*	
	Published	*Adjusted*	*Published*	*Adjusted*	*Published*	*Adjusted*	*Published*	*Adjusted*	*Published*	*Adjusted*	*Published*	*Adjusted*	*Published*	*Adjusted*
Estimated number of households in population ('000)	397.8	456.9	456.3	413.9	572.9	521.7	510.0	475.4	378.3	384.9	319.4	381.9	2,634.7	2634.7
Current housing costs	16.0	16.1	16.2	15.8	15.8	16.4	15.4	15.3	13.6	14.0	11.6	12.1	14.5	14.6
Fuel and power	4.0	3.4	2.7	2.5	2.4	2.5	2.1	2.1	2.1	2.0	1.7	1.7	2.3	2.2
Food	26.0	24.6	22.5	22.4	21.6	22.2	20.5	20.8	19.6	19.9	18.2	18.6	20.6	20.9
Alcohol and tobacco	5.1	5.2	5.9	5.6	5.7	5.9	5.9	5.8	6.0	6.1	6.0	6.1	5.9	5.9
Clothing and footwear	8.3	8.3	7.8	8.3	8.2	8.2	8.6	8.6	8.9	9.5	10.4	10.2	8.9	9.0
Household equipment and operation	10.0	8.6	9.0	10.4	9.5	9.1	8.8	9.3	9.3	9.1	9.7	10.1	9.3	9.5
Medical care and health expenses	3.6	3.2	4.2	3.3	4.0	3.4	3.8	3.2	3.6	2.9	3.2	2.4	3.7	3.0
Transport and communication	12.5	13.9	15.9	15.7	16.3	16.3	17.6	17.8	17.9	17.9	16.9	17.1	16.7	16.8
Recreation and education	5.8	7.3	7.5	7.4	7.8	7.8	8.2	7.9	9.4	8.6	11.3	11.4	8.8	8.8
Miscellaneous goods and services	8.6	9.3	8.3	8.7	8.6	8.2	8.9	9.1	9.6	10.1	10.9	10.2	9.3	9.4
Total expenditure	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note: Discrepancies between the sum of components and totals are due to rounding.

There were three main sources of data against which the survey data could be compared. These are discussed below and some examples of the comparisons are given.

(i) Australian National Accounts (ABS publication Catalogue No. 5204.0 Previously Ref. No. 7.1): This was the major data source for comparison with the Household Expenditure Survey expenditure figures. The major differences between the two sets of figures were:

- National Accounts figures included expenditure by non–profit organisations servicing the households.
- National Accounts figures are based on the value of acquisitions of households rather than payments made by households.
- National Accounts estimates are for the whole of Australia whereas the Household Expenditure Survey gave estimates for capital cities only.
- National Accounts commodity groupings were different from the broad expenditure groups used in the survey. However, it was possible to adjust for this regrouping.

Comparisons made at the broad level showed that there were few discrepancies which were larger than those expected because of the differences in the two collections. Items which showed unexplained differences were 'Alcohol and tobacco', 'Books, newspapers and magazines', 'Fares' and 'Entertainment and recreational services'. In each of these cases the Household Expenditure Survey estimate understated the National Accounts figure.

The table below compares some of the estimates derived from National Accounts with the Household Expenditure Survey estimates.

NATIONAL ACCOUNTS AND HES EXPENDITURE ESTIMATES

Category of Expenditure	*Weekly Household Expenditure Estimate ($)*	
	National Accounts	*HES*
Food	28.10	32.38
Alcohol	9.11	5.92
Cigarettes and tobacco	4.11	3.33
Footwear	1.98	2.34
Purchase of motor vehicles	8.10	7.41
Fuel & power	3.47	3.54
Postal, telephone and telegraphic charges	2.08	1.99
Fares	5.01	2.60

(ii) 1971 Population Census (ABS publications Reference Nos. 2.83 to 2.97): Figures from the 1971 Population Census were used to verify the survey estimates of numbers of households and persons. Because the Census collected details of the characteristics of households and persons (eg occupation of head, nature of housing occupancy) which are used in classifying households in the Household Expenditure Survey results, it was possible to use this data to check the distribution of sample households according to these characteristics.

The major difficulty in using Census data was allowing for changes in population size and composition which occurred in the period between the Census and the survey. For the purposes of checking most distributions it was assumed that although absolute levels of population had changed, the proportion of the population in each class would remain almost constant.

The Population Census and Household Expenditure Survey demonstrated similar results where comparison was possible. Examples of this comparison for occupation of head and nature of housing occupancy are illustrated in the following tables:

CENSUS AND HES ESTIMATES

Occupation of Head	Percentage of Total	
	Census (a)	HES (b)
Professional and Administrative	23	25
Clerical and Sales	16	18
Tradesmen and Labourers	42	41
Other	19	16

(a) Includes both employees and self-employed persons who are head of dwelling. (b) Employee head of household only.

Nature of Housing Occupancy	Percentage of Total	
	Census (a)	HES (b)
Renting	28	29
Owner–Occupied, Process of Purchase	72	71

(a) Refers to dwellings (b) Excludes other and not stated

(iii) 1973–74 Supplementary Survey of Income (ABS publication Reference No.17.6): This survey provided the most current income data which could be used for comparison with the Household Expenditure Survey income figures. However, comparison was particularly difficult because:

- During the period between the two surveys, marked increases in income occurred, with Average Weekly Earnings per employed male unit rising 28% from December 1973 to December 1974. The rise in incomes from sources other than wages and salaries were not as marked as this and therefore it was inexact to apply a blanket rate to adjust the data.
- The Survey of Income included persons in non–private dwellings while the Household Expenditure Survey did not.
- Answers to the Survey of Income questions were frequently provided by one person from the household and were based on memory rather than records.

Thus the comparison had to be at a very broad level. The table below compares results from the two surveys for the six standard income ranges used in most Household Expenditure Survey publications.

INCOME SURVEY AND HES ESTIMATES

Weekly Household Income	Percentage of Total Households	
	1973–74 Supplementary Income Survey (a)	HES
Under $80	17	15
$80 and under $140	15	17
$140 and under $200	22	22
$200 and under $260	18	20
$260 and under $340	15	14
$340 and more	13	12

(a) Income from this survey was aggregated to the same household unit used in the Household Expenditure Survey and estimates were derived for the six State capital cities and Canberra. An adjustment of 25 per cent was applied to all household income in order to make allowance for increases in income between 1973–74 and 1974–75.

Further Information

For more details of any of the studies mentioned here or any other aspects of the non-sampling errors associated with this bulletin please contact Mr Peter Gardner, Canberra 525609.

APPENDIX 2 –STANDARD ERROR GRAPHING PROCEDURES

The estimation formulae given in Appendix 4 are quite complex because of the complexity of the sample design and the need to adjust for non–response. Consequently explicit variance and standard error estimation is also quite complex. For this reason a replicated variance estimator (known as split fourths) was used in the estimation of standard errors for the survey. Full details of the split fourths variance formulae are given in Appendix 4. Because the split fourths variance estimator can be expected to be less stable than the explicit estimator, and because it permitted a neater presentation at smaller cost, standard errors were 'smoothed' where possible and presented in graphical form. The rationale behind the models used in this smoothing process is discussed below.

If the survey had been based on a simple random sample without replacement the standard error percent of an estimate would be given by

$$SE\% = \frac{V}{\sqrt{n}} \times 100$$

where V is the coefficient of variation of the item being measured and n is the sample size.

The effect of the complex sample design can be assumed to influence the standard error through a constant K as follows

$$SE\% = K\frac{V}{\sqrt{n}} \times 100$$

where K is a constant known as the design effect reflecting the effects of the complex sample design; eg if the design is less efficient than a simple random sample without replacement, K will be greater than 1.

Taking logs gives an expression of the form

$$\log SE\% = \log(100\,KV) - \tfrac{1}{2}\log n$$

Note that the coefficient of variation of an estimate is a measure of the population variability of the item being measured relative to its mean, and depends on the following factors:

(a) the item itself; eg expenditure on 'Electricity' would be expected to have a smaller coefficient of variation than expenditure on 'Household Appliances', since it is a less variable item.

(b) the way the item is classified; ie the coefficient of variation for a particular expenditure item may be expected to vary between different income ranges, capital cities, quarters etc.

The design effect may also depend on the item being measured and the way it is classified.

The model actually used was

$$\log(SE\%) = a + b \log n$$

In this model 'b' can be expected to be fairly close to –½ and 'a' depends on 'V' and 'K'. A different relationship may be applicable to each item and each classification. However, evidence indicated that the difference between items was more important than differences between classification variables and ranges, capital cities, quarters etc. A different relationship was therefore fitted to each item using selected standard error percents from different classifications to obtain the ordinary least squares regression estimates for 'a' and 'b'. The resultant regression lines were plotted on log – log scale to form the standard error graphs as presented in this bulletin. In some cases items were combined where their lines were similar.

The 'R^2' values, which provide an indication of the strength of the linear association between the standard error percent and the sample size, were typically between 0.6 and 0.9 indicating reasonable fits to the model. The estimate 'b' was generally between –0.4 and –0.5. The apparent flatness of the lines may have been the result of smaller coefficients of variation for estimates based on small sample sizes than for those based on large sample sizes. This is because the smaller groups of households (or persons) tended to be more homogeneous, and thus had smaller population variability than larger groups (eg. expenditure items would be subject to larger variability across income groups than within income groups). This also explains why some standard error lines cross. For example on Graph I(13), expenditure on 'Tobacco' may be less elastic (eg. with respect to income) than expenditure on 'Wine' and therefore the 'Wine' curve is flatter than the 'Tobacco' curve. The high 'R^2' values also suggest that the split fourths variance estimator did in fact provide reasonably stable variance estimates.

In some cases for numbers of persons or numbers of households estimates, standard error percents were plotted against the estimates themselves. In the case of an estimate of number of persons or households with a particular characteristic

$$V \cong \frac{1}{\sqrt{p}}$$

where p is the estimated percentage of persons or households with that characteristic, and p is small.

Thus

$$SE\% = K \frac{1}{\sqrt{np}} \times 100$$

$$= K \frac{\sqrt{N}}{\sqrt{n}} \quad \frac{1}{\sqrt{Np}} \times 100$$

where N is the population number and Np is therefore the estimate.

Thus taking logs

$$\log (SE\%) = \log (100K \frac{\sqrt{N}}{\sqrt{n}}) - \tfrac{1}{2} \log Np$$

and therefore a model of the form

$$\log (SE\%) = a + b \log (\text{estimate})$$

is appropriate.

APPENDIX 3 -- SAMPLE DESIGN AND SELECTION

SAMPLE DESIGN

Development of an efficient sample design is basically concerned with deciding on an appropriate sample size, and how the sample is to be distributed between and within areas. This entails consideration of factors such as the topics covered, the level and accuracy at which the statistics are required, the cost of conducting the survey, and any operational constraints. For the 1974-75 HES it was decided that a sample of about 9,100 fully cooperative households would be required, distributed among the six State capital cities and Canberra. Darwin was initially included in the survey but data collection was abandoned in that city following cyclone Tracy in December 1974.

For this survey households were selected by the process of first selecting private dwellings and then identifying the households within each of these dwellings. In determining the total sample size (i.e. the number of dwellings to be selected to yield the required number of cooperating households) allowance had to be made for vacant dwellings, households that could not be contacted, households found to be outside the scope of the survey, and households unwilling or unable to provide all the required information. The following table shows the required number of fully cooperating households for each capital city and Canberra and the number of dwellings selected in anticipation of obtaining that number of cooperating households.

SAMPLE SIZE

City	*Required Number of Cooperating Households*	*Total Sample Size (No. of dwellings selected)*
Sydney	2400	3800
Melbourne	2400	3700
Brisbane	1100	1700
Adelaide	1050	1400
Perth	950	1400
Hobart	600	750
Canberra	600	800
Total	**9100**	**13550**

The sample was, initially, optimally allocated for the production of estimates for the seven cities combined, and then adjusted to achieve acceptable standard errors for the estimates for each capital city, particularly as separate capital city estimates were required for purposes of reweighting the Consumer Price Index. This adjustment has implied a slight reduction in the accuracy of estimates for the seven cities combined, compared with what would have been possible from an unadjusted optimal sample allocation.

In order to improve the efficiency of the sample and hence the accuracy of the estimates, each capital city was divided into a number of relatively homogeneous areas or strata . A stratum consisted of a Local Government Area or a group of contiguous Local Government Areas and was determined in such a way that households within the stratum would be as alike as possible with respect to selected socio-economic characteristics. Within each city the sample was allocated to strata in such a way as to increase the representation of pensioner, migrant, and low income households in the sample, while maintaining adequate representation of other groups. This was done to improve the accuracy of estimates for these special groups. The increased representation was adjusted for in the estimation process to provide unbiased estimates for all households.

SAMPLE SELECTION

Within each stratum the sample was selected in three stages. Each stratum contained a number of Census Collector's Districts (CDs), which had been determined for the 1971 Census of Population and Housing (a CD generally contains about 250 dwellings). A sample of CDs was selected from the stratum with probability proportional to number of dwellings (2,300 CDs in total were selected). Proportionally more CDs were selected from those strata known to contain a large proportion of special groups such as pensioners and migrants.

Each selected CD was divided into blocks which consisted of between 20 and 40 dwellings. Two blocks were selected from each selected CD with each block having a probability of selection proportional to the number of dwellings it contained. Finally, within each selected block, lists of all dwellings defined as being within the scope of the survey were prepared and a systematic random sample of dwellings was selected. This process of selection ensured that every dwelling within a stratum had the same chance of selection.

In order to take account of possible seasonal effects the sample was allocated equally to each quarter of the year. Initially CDs were allocated to quarters on a random basis. One block was taken in the quarter to which the CD was initially allocated. For CDs allocated to the September, December and March quarters, the second block was taken in the following quarter i.e. December, March and June quarters respectively. For CDs allocated to the June quarter, the block was taken in the previous September quarter. Once a block was allocated to a quarter, the month for data collection was determined at random, whilst ensuring that the same number of blocks was allocated to each month of the quarter. Within a month, one half of the sample was enumerated in each half of the month. This selection process ensured an even spread of dwellings throughout the year, and enabled quarterly estimates as well as annual estimates to be obtained from the survey.

APPENDIX 4 - ESTIMATION AND VARIANCE ESTIMATION PROCEDURES

1 ESTIMATES OF LEVEL

(a) Estimation

The estimate of level for a variable X is given by

$$x'' = \sum_q \sum_h \sum_a \sum_j w_{qha}\, x_{qhaj}$$

where $$w_{qha} = \frac{1}{f_h} \times \frac{n_{qh}}{1n_{qh} + 2N_{qh}} \times \frac{1n_{qha} + 2n_{qha}}{1N_{qha}}$$

where q sums over quarters

h sums over strata

a sums over post-strata 1, 2, 3

(See Sections 5 and 8.6 of Bulletin 1 for descriptions of strata and post-strata)

j sums over responding households

f_h = sampling fraction in stratum h

n_{qh} = total sample size in quarter q, stratum h (responding and non-responding households)

$1n_{qh}$ = number of responding households in sample in stratum h, quarter q

$1n_{qha}$ = number of responding households in post-stratum a, stratum h, quarter q

$2n_{qha}$ = number of non-responding households in post-stratum a, stratum h, quarter q

$2N_{qh} = \sum_a 2N_{qha}$

where $$2N_{qha} = \begin{cases} 0 & \text{if } 1n_{qha} = 0 \\ 2n_{qha} & \text{otherwise} \end{cases}$$

$$1N_{qha} = \begin{cases} 1 & \text{if } 1n_{qha} = 0 \\ 1n_{qha} & \text{otherwise} \end{cases}$$

x_{qhaj} = value of item X for household j, post-stratum a, stratum h, quarter q. For example, for "number of household" estimates x_{qhaj} will be a (1,0) variable according to whether the household has the characteristic or not.

Basically the estimation procedure consists of three stages.

(i) Assigning the simple stratum raising factor $1/f_h$ to each record in the stratum.

(ii) An adjustment for non-respondents who can be classified into one of the three "type and size" post-strata containing at least one respondent. This involves the factor

$$\frac{1^n_{qha} + 2^n_{qha}}{1^N_{qha}}$$

(iii) An adjustment for non-respondents who cannot be classified into post-strata or who are allocated to a post-stratum with no respondents. This involves the factor

$$\frac{n_{qh}}{1^n_{qh} + 2^N_{qh}}$$

w_{qha} is the survey "weight" or "expansion factor" to be applied to all respondent households in quarter q, stratum h and post-stratum a.

(b) Variance estimates for levels

Variance estimates are obtained by collapsing the original strata to 14 broad strata, and applying the split-fourths technique at this level. Thus the variance estimator for x" is

$$\hat{\text{var}}(x") \doteq \frac{1}{k(k-1)} \sum_{b=1}^{14} \sum_{i=1}^{k} \left[kx'_{bi} - x'_b \right]^2$$

where the sample is split into k = 4 sub-samples called variance groups.

i sums over variance groups

b sums over broad strata

$$x'_b = \sum_q \sum_{h \in b} \sum_a \sum_i \sum_j w_{qha} x_{qhaij}$$

$$x'_{bi} = \sum_q \sum_{h \in b} \sum_a \sum_j w_{qha} x_{qhaij}$$

where q sums over quarters

a sums over post-strata

j sums over responding households

x_{qhaij} = value of item X for household j, variance group i, post-stratum a, stratum h, quarter q.

The estimated standard error percentages for estimates of level are given by

$$SE\% = \frac{\sqrt{(\hat{var}(x''))}}{x''} \cdot 100\%$$

2 ESTIMATES OF RATIO

(a) Estimation

The estimate of the ratio R of the level of variable X to the level of variable Y is given by

$$r = \frac{x''}{y''}$$

where x", y" are estimates of level as given above. For estimates such as average expenditure for households with a particular characteristic, x" is the estimate of the level of total expenditure by households with that characteristic, and y" is the estimated number of households in the population with that characteristic.

(b) Variance estimates for ratios

The split-fourths variance estimator for estimates of ratios is given by

$$\hat{var}(r) = \frac{1}{k(k-1)} \frac{1}{(y'')^2} \sum_b \sum_i \left[(kx'_{bi} - x'_b) - r(ky'_{bi} - y'_b) \right]^2$$

where k, b, i, x'_b and x'_{bi} are as in section 1(b) and y'_b, y'_{bi} are defined similarly to x'_b, x'_{bi}.

The estimated standard error percentages for estimates of ratio are given by

$$SE\% = \frac{\sqrt{\hat{var}(r)}}{r} \cdot 100\%$$

3 ESTIMATES OF PERCENTILES

A percentile of a variable X associated with a percentage p is a point X_p in the range of X values such that p percent of the population have X values less than X_p. For example the median of a variable X (i.e. p = 50%) is the point X_{50} such that 50% of the population has X values less than X_{50}.

(a) Estimation

Percentiles are estimated as follows:

The range of values for the variable X is divided into a number of classes, and the population numbers R_m (m = 1, ...M) falling into the classes are estimated by summing the weights W_{qha} for the responding households in each class; i.e. a weighted sample histogram is formed. Suppose the m'th class with class boundaries (X_{m-1}, X_m) contains the point below which an estimated p% of the population falls. The estimate of the p'th percentile is given by

$$x'_p = X_{m-1} + \frac{\frac{p}{100} \sum_{i=1}^{M} R_i - \sum_{i=1}^{m-1} R_i}{\sum_{i=1}^{m} R_i - \sum_{i=1}^{m-1} R_i} \left[X_m - X_{m-1}\right]$$

(b) Variance estimates for percentiles

The estimate of the variance of x'_p is given by

$$\hat{var}(x'_p) = C \cdot \frac{\frac{p}{100}\left(1 - \frac{p}{100}\right)}{n\, f_p^2}$$

where n is the sample size

f_p is the estimated proportion of population per unit interval in the class around x'_p.

i.e.
$$f_p = \frac{R_m}{(X_m - X_{m-1}) \sum_{i=1}^{M} R_i}$$

C is a cluster factor to take account of the fact that the estimate was based on a multistage sample design. In this case C = 2 has been chosen.

1974–75 HOUSEHOLD EXPENDITURE SURVEY PUBLICATIONS

The results of the 1974–75 Household Expenditure Survey will be issued progressively in a series of eight bulletins. The bulletins (listed below) may be purchased through the mail from Mail Order Sales, Australian Government Publishing Service, P.O. Box 84, Canberra, A.C.T. 2600; over the counter from the Government Publications and Inquiry Centres in each State capital; and through the mail or over the counter from offices of the Australian Bureau of Statistics in each State capital.

The title and reference number of each publication is shown below, followed by its price, the price including postage, and a brief summary of contents.

BULLETIN 1 – AN OUTLINE OF CONCEPTS, METHODOLOGY AND PROCEDURES (Catalogue No. 6507.0 Previously Ref.No. 17.19) ($1.20, $1.60)

Provides a detailed description of survey objectives, scope and coverage, concepts and definitions used, and sample design and methodology.

BULLETIN 2 – PRELIMINARY RESULTS (Catalogue No. 6508.0 Previously Ref. No. 17.20) (70c, $1.00)

Contains, for each capital city, estimates of household expenditure (broad groupings) classified by income. Also contains, for all capital cities combined, estimates of household expenditure (broad groupings) classified by household composition, work status and occupation of the household head, age of household head, and estimates for each quarter of 1974–75.

BULLETIN 3 – STANDARD ERRORS (Catalogue No. 6509.0 Previously Ref. No. 17.21) (80c, $1.20)

Contains estimates of sample errors associated with data items in the survey, for use in assessing the reliability of particular survey results.

BULLETIN 4 – EXPENDITURE CLASSIFIED BY INCOME OF HOUSEHOLD (Catalogue No. 6510.0 Previously Ref. No. 17.22) ($1.10, $1.50)

Contains estimates of household expenditure (detailed groupings) classified by income, for each capital city.

BULLETIN 5 – QUARTERLY EXPENDITURE PATTERNS (Catalogue No. 6511.0 Previously Ref. No. 17.23) (60c, 90c)

Contains estimates of household expenditure for each quarter of 1974–75, for each capital city.

BULLETIN 6 – EXPENDITURE CLASSIFIED BY HOUSEHOLD COMPOSITION (Catalogue No. 6512.0 Previously Ref. No. 17.24) (90c, $1.30)

Contains estimates of household expenditure by income and household composition (that is, number of adults and children).

BULLETIN 7 – INCOME DISTRIBUTION (Catalogue No. 6513.0 Previously Ref. No. 17.25) ($1.00, $1.40)

Contains information on the sources of income and distribution of income for households and for persons.

BULLETIN 8 – EXPENDITURE CLASSIFIED BY SELECTED HOUSEHOLD CHARACTERISTICS (Catalogue No. 6514.0 Previously Ref. No. 17.26) ($1.10, $1.50)

Contains expenditure estimates for households with selected characteristics (for example, where the head of household was born overseas; pensioner households).

Further details of the tables included in each of the above publications can be found in Bulletin 1.

Apart from the published statistics it is intended to provide some additional tabulations of data in response to requests from users. There is little possibility of meeting such requests immediately because of heavy computer programming commitments in connection with publications and certain other urgent requirements which have to be met. The confidentiality requirements of the Census and Statistics Act will also impose limitations on the amount of detail that can be provided. Requests for special tabulations may be directed to the Australian Bureau of Statistics, P.O. Box 10, Belconnen, A.C.T. 2616. Such requests will be acknowledged and will be considered as soon as resources permit.

(R 77/543) Cat. No. 77 6900 4

24049/78—L